Against Conditionalization

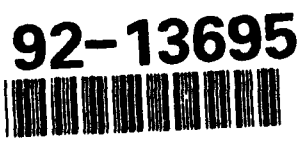
Bacchus, Kyburg, Thalos

March 16, 1989

1 Introduction

Bayesian epistemology is marked by a scruple for compliance with the probability axioms. One cornerstone of Bayesian epistemology is the doctrine of personalism, the view according to which an agent's beliefs are not the mechanical result of conditionalizing a logical probability over her total history of observational experience. Another cornerstone of Bayesian epistemology is the teaching that since personalism is true, epistemic injunctions must be issued to rational agents to procure their compliance with the probability axioms, so that their beliefs are characterized by real-valued degrees that are *coherent* in the technical sense of being governed by the same constraints that rightly rule measures of objective chance. As a result Bayesians brandish Dutch Book theorems, tout conditionalization as the only true path to new beliefs in responce to new evidence, and endorse the principle of Reflection as the price of personal epistemic integrity.

In this paper we argue that the epistemic levies which Bayesians exact in return for bestowing the benison of rationality on human believers are extortionate. We propose to pose a systematic challenge to Bayesian principles, from Dutch Book to conditionalization to Reflection, focusing on the issue of conditionalization. We will show that conditionalization is by no means the only rational method of updating belief (if it is a rational method at all). And the reasons we will delineate in favor of this view will cast doubt on both Dutch Book arguments and Reflection. We will show that an agent might and sometimes ought be counted rational even if he does not conditionalize or Reflect or avow Dutch Book. These principles, we will demonstrate, dis-



REPORT DOCUMENTATION PAGE

Form Approved
OPM No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1989	Unknown

4. TITLE AND SUBTITLE

Aganinst Conditionalization

5. FUNDING NUMBERS

DAAB10-86-C-0567

6. AUTHOR(S)

Bacchus/Kyburg/Thalos

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Rochester
Department of Philosophy
Rochester, NY 14627

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army CECOM Signals Warfare Directorate
Vint Hill Farms Staiton
Warrenton, VA 22186-5100

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

92-TRF-0017

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Statement A; Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

In this paper we argue that the epistemic levies which Bayesians exact in return for bestowing the benison of rationality on human believers are extortionate. We propose to pose a systematic challenge to Bayesian principles, from Dutch Book to conditionalization to Reflection, focusing on the issue of conditionalization. We will show that conditionalization is by no means the only rational method of updating belief (if it is a rational method at all). And the reasons we will delinate in favor of this view will cast doubt on both Dutch Book argument and Reflection. We will show that an agent might and sometimes ought be counted rational even if does not conditionalize or Reflect or avow Dutch Book. These principle, we will demonstrate, discount too much which is rational as unworthy. We will cry "Justice!" and proclaim that rationality need not come as dear as they insist.

14. SUBJECT TERMS

Artificial Ingelligence, Data Fusion, Conditionalization

15. NUMBER OF PAGES

33

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UCNLASSIFIED	UL

count too much which is rational as unworthy. We will cry "Justice!" and proclaim that rationality need not come as dear as they insist.

We begin first by investigating and spelling out what is required to establish the epistemic imperatives which constitute the conclusions of Dutch Book arguments. We will argue that many of the premises required are highly suspect from an intuitive point of view. We will then turn our attention to efforts to justify updating by conditionalization. We focus on conditionalization because it affords the clearest spectacle of the Bayesian perspective on belief: how the Bayesian regards the human believer is made most manifest in how the Bayesian constrains the believer to change belief in light of new evidence. We will argue that the attempts to justify conditionalization fail and conclude that the view of the human believer which is implicit in the Bayesian cluster of principles is simply mistaken.

2 Static dutch book arguments and conditional bets

Reduced to its bare essentials, the dutch book argument for static beliefs that measure up to the classical probability calculus—leaving aside such niceties as strict coherence, conglomerability, and the like—goes as follows: If your degrees of belief don't satisfy the axioms of the probability calculus, you can have a dutch book made against you, according to which you will lose no matter what happens.

In view of the fact that this claim is sometimes referred to as "The Dutch Book Theorem," we may suspect that there is more to it than a matter of bare assertion. On the other hand, the premises required to derive the practical import of the conclusion are rarely spelled out in full.

In the first place, as has been pointed out by Kyburg [1], Chihara and Kennedy [2], Baillie [3], and Schick [4], and no doubt others, the "agent" is not going to have a book made against him unless he accepts a set of wagers according to which he loses no matter what happens. But that he should not accept such a set of wagers, if he would prefer not to lose no matter what happens, is a matter of deductive logic, and has nothing to do with what *degrees* of belief he may have, if any. That I am bound to lose a dollar if I bet on heads at odds of two dollars to one, and also on tails at odds of two

dollars to one, has nothing to do with my degrees of belief, nor with whether or not the coin in question is fair. It is simply a deductive consequence of the fact that in every world we regard as possible, either heads and no tails or tails and no heads represents the result of the coin toss. One need not invoke the probability calculus in order to enjoin a rational agent from committing himself to a sure loss of what he values!

So what is the supposition involved here? One supposition that would make sense of the Dutch Book Principle is that a rational agent would be willing to take either side of a bet on any proposition at odds corresponding to his degree of belief. That is, if his degree of belief is p, the odds he would offer to bet at are p to $1-p$ *on* the proposition, and odds of $1-p$ to p *against* it. *If* this were true, then it would follow that the degrees of belief of the agent in related propositions would have to satisfy the constraints imposed by the probability calculus.

But this is surely not true. There are classical worries about people who love to gamble, and will pay a premium for the priviledge of taking a risk, and about people who are upset by uncertainty, and will pay a premium not to gamble. We leave those to one side here. Consider only a perfectly cold-blooded and rational man, who neither suffers anxiety nor gets excitement from betting. All he is concerned about is the money.[1] Even this individual, however, will refuse to make bets at odds determined by his degrees of belief (if any).

The reason is that there is some lapse of time between the time that a bet is placed and the time that it is settled. Suppose that the agent has a degree of belief equal to p in the statement S. He is willing to bet at odds of p: $1-p$ on S, and at odds of $1-p$: p against S, according to the principle in question. But to make both bets for unit stakes is to tie up one unit of utility until it is determined whether or not S is true. During that interval the rational man will want a return on his committed capital; he will expect a return to compensate him for the use of the money involved. Thus the cold-blooded agent, to whom gambling is neither attractive nor repulsive, will still want compensation for the use of his capital. This translates directly, on the assumption of the usual relation between degrees of belief and odds, into the

[1] Here and elsewhere we refer to money, to dollars, and so on, when we should more properly refer to units of utility in order to take account of the (eventually) decreasing marginal utility of money and allied phenomena.

requirement that the degrees of belief of the rational agent in S and $\neg S$ *must* add up to *less* than 1.0.

Is this just a small matter of idealization? In celestial mechanics, after all, we suppose that the planets are point masses. But two senses of idealization are involved: descriptive and normative. In the former sense, we could easily forgive the fact that the degrees of the rational agent should add up to less than 1.0, how much less depending on the date of the settlement of the bet. But in the latter sense—the sense that we take to be of concern in epistemology—this fact is important. Should one merely make sure that one's *actual* bets not lead one to a sure loss (for which deductive logic is perfectly sufficient) or should one be concerned about hypothetical bets?

What is required to derive the dutch book principle is a much stronger premise (correctly noted by Anscombe and Aumann [5])—namely that the agent must be *compelled* to post odds on the set of propositions at issue, and *compelled* to take all bets offerred at these odds. Under these circumstances—under which the agent is not allowed a fair return on his capital—it is indeed true that the agent must post odds corresponding to some coherent set of probabilities—i.e. a set of probabilities satisfying the axioms of the probability calculus.

But again we have lost the connection to degrees of belief. No matter what the degrees of belief (if any) of the rational agent, no matter what odds he would be willing to offer on any particular bet, it is a matter of insurance against the worst case that he should post odds that correspond to probabilities satisfying the classical calculus.

Insurance against the worst case? That suggests that there are other cases. And that requires another doubtful premise. It is true that if the agent is compelled to post odds, and is compelled to take any bet at those odds, the only way he can protect himself against the possibility of certain loss is by posting odds that correspond to a coherent set of probabilities. But this corresponds to the worry that there is a very smart bettor out there, trying to take advantage of him, whose utilities correspond in important ways exactly to the utilities of the agent.

Why should we suppose that the world is thus uncooperative? Just because it is *possible* that a book should be made against the agent does not mean that a book *will* be made against the agent. And if it is possible that no book is made against him, there is no *need* for him to lose under all circumstances. For the modal argument to go through leading to the conclusion

that the agent *must* post coherent odds, we need a non-trivial existential assumption. The weaker conclusion, that if the agent posts incoherent odds it is *possible* that he could have book made against him, is hard to distnighish pragmatically from the assertion that on any finite set of bets—at whatever odds—the agent could lose (unless he is the lucky bookie).

Now it may be that it is a principle of rationality that if you are compelled to post odds on a set of statements, and compelled to take all bets at those odds (presumably in units of your utilities), then it is only rational to be so suspicious of the world that you should not allow the *possibility* of being taken advantage of by an evil and intelligent bettor. We ourselves do not find this completely persuasive: it seems to smack more of paranoia than rationality.

Be that as it may, there is still the question of why the odds posted should reflect the agent's degrees of belief. Why should it not be the case that the agent has a set of degrees of belief, and at the same time posts odds that would correspond to a different set of degrees? At the most superficial level, one may simply say that these odds represent what degrees of belief *are*.

A somewhat deeper answer is that an agent's *expectation*, calculated in terms of his degrees of belief, would be negative. This requires unpacking, since it (again) depends on facts about the world. Suppose I am offered exactly one bet, at even money, on tails. I accept it. I have a degree of belief of .4 that I will win, of .6 that I will lose. I am certainly not *assured* of loss. Let us suppose that I am offered, and am compelled to accept, a large finite number of bets concerning the next toss of this coin, or concerning a sequence of tosses of this coin that I suppose to be characterized in the same way. In any finite set of bets at even money on heads and even money on tails, only three things can happen, regardless of my degrees of belief: I will come out ahead; I will come out behind; or I will break even. All three remain possibilities. Given that the odds I post satisfy the constraints imposed by the probability calculus, however, I can be sure that there is no possibility that I will be made to take a set of bets under which I will lose no matter what happens. This does *not* mean that I won't lose; only that I won't be dutch booked.

Suppose that there are rational degrees of belief. Suppose that we have a meter that measures the actual degree of belief of an agent in a proposition

S.[2] Suppose also that the agent, a full convert to Dutch book, is compelled to post odds. Then it will be the case that the odds posted by the agent under the circumstances outlined satisfy the probability calculus, but it may or may not be the case that the *rational degrees of belief* of the agent will also conform to the probability calculus.

The reason is that Dutch book considerations bear only on the rectitude of the coherence of the *odds posted*, but they have no direct bearing on the rectitude of *belief*. What is required to constrain belief is something over and above Dutch book consideration. The following principle comes to mind as a principle of the sort needed to do the proper job: agents must believe in consonance with (according to) the odds that they post. But this principle falls so far short of plausiblity as to verge on nonsense.

The foregoing principle is confused with the following more plausible principle: Agents must post odds in accord with their beliefs. But this principle does not yield the Bayesian result because nothing in the Dutch book argument itself applies to *belief:* belief has nothing to do with losing money come what may— only betting badly does. So if Dutch book arguments don't apply to belief, then invoking the more plausible principle does not help; it is irrelevant.

Finally, there is the question of what rationality dictates in the case of an agent who is constrained to post odds, to comply with Dutch book, to take all bets at the odds posted, and to believe in accordance with the posted odds. We are persuaded that rationality ordains nothing (beyond deductive constraints) in this unfortunate agent's case. He must be guided by the light of prudence. Even among the alternatives permitted by the constraints there are a multitude of rationally acceptable ones. (For example, an agent might post odds on heads on a given toss of a coin anywhere between 45:55 and 55:45, and believe accordingly.)

There is an argument for the identity of degrees of belief and propensities to bet. It is the behavioristic argument that the only way to *measure* the agent's degrees of belief (rational or otherwise) is by means of the odds that we have compelled him to post. But this argument is not a compelling argument. It is only as persuasive as the general argument for behaviorism.

[2]Here is a handy use for robots and artificial intelligence; when we can make rational robots, we can no doubt provide a meter that indicates the degree of belief of the rational robot in any given proposition. So we can input the proposition and the robot will output its rational degree of belief.

Nay, worse, for constrained behavior may not be as revealing as unconstrained behavior.

We have so far left to one side another assumption of the dutch book argument, except for some subtle parentheses. This is the assumption that there *are* "degrees" of belief. One certainly does not arrive at the idea that one's degree of belief in S is measured by a real number in the closed interval [0,1] by introspection. My feeling about rain tomorrow, at any rate, does not correspond to any real number. Of course I can be compelled—just as I can be compelled to post odds on rain—to name a price that I would either pay for a ticket that would pay a dollar in the event of rain, or that I would sell a ticket for that I would redeem for a dollar in the event of rain. But this concerns the impact of compulsion, not the character of my psychological states. If I merely introspect, I do not find a real number.

What else might one find? An interval? Savage rightly observed that it is at least twice as hard (actually exponentially as hard) to determine the bounds of intervals as it is to determine real-number degrees of belief. But this is a red herring. In a given situation, it is certainly easier to specify odds against S that the agent will not accept, and odds against $\neg S$ that the agent will not accept, than it is to find a set of odds such that the agent will be willing to accept either side of a bet at those odds. (As indicated earlier, an agent who takes the value of his money over time seriously will deny that there is any such set of odds!)

So it is really quite easy to *limit* the range of odds that are acceptable. Intervals, as measures of rational belief, seem to be not so implausible after all.

Let us suppose that there is a set of intervals characterizing the belief in each of a set of propositions. We might get at them by the mild behavioristic method of determining acceptable odds, or by some other method. We may ask what the relation is between this set of intervals and the odds that the agent might be compelled to post for all bettors. In particular, is there any argument to the effect that the intervals characterizing the rational agent's degrees of belief must *allow* a set of odds to be posted that conform to them. Put more precisely: is it the case that there must be a belief function from statements to real numbers that satisfies both the calculus of probability and the constraints imposed by the belief intervals of an agent. We know of no such argument. The compulsorily posted odds must correspond to a belief function that satisfies the calculus of probability. As already observed, that

is a deductive constraint that has nothing to do with degrees of belief. That this belief function also satisfies the interval constraints reflecting the agent's doxastic state seems desirable, intuitively, but it is not at all clear that there is a persuasive argument that this *must* be the case.

In point of fact, it is the case for epistemological probability (Kyburg [6]) that for any finite field of propositions there exists a belief function satisfying the axioms of probability whose values fall in the epistemological probability intervals. This is a nice feature of that system: it says that you need not depart from rationality in order to insure yourself against the possibility of being dutch booked.

Nevertheless, one might want to forgo the insurance. If nobody is out to get you, you don't have to look in all the dark corners. And what is the advantage of *not* looking in all the dark corners? Why, to take advantage of what happens almost all the time. Suppose that your interval of degrees of belief in S is [0.3,0.4], but almost nobody bets against it. Then you might publicly offer odds of even money for bets on S, since you can be practically certain that you will have to cover very few bets against S compared to the number of bets you will get on S.

To sum up so far: The Dutch Book Principle presupposes that there are (real-valued) degrees of belief; that the agent should have degrees of belief that correspond to the odds that he is willing to post on various propositions; that he is committed to taking all bets at these odds on these propositions; that there is an active conspiracy on the part of his takers to take advantage of him; and that rationality consists in guarding against this possibility.

This is a lot to swallow, but many highly respectable philosophers have managed it. There is, however, a set of consequences of these principles that still needs to be spelled out—namely, the consequences concerning "conditional" or "called off" bets.

2.1 Conditional Bets

A conditional bet on S given T, at odds of $r{:}\,p$ is a bet in which the better agrees to receive from the bookie r units if S and T are both true, to pay the bookie p units if $\neg S$ and T are both true, and to call the whole bet off if T happens to be false. Assuming that the dutch book principle applies, and *in addition* that the agent must post odds on conditional bets as well as straight bets, it can be shown (first by Frank Ramsey [7]) that the probability

(determining the odds) of a conditional statement (S given T) must be the ratio of the probability of S and T to the probability of T. That is, if my conditional degree of belief in S given T is to determine betting odds consonant with my degrees of belief in S and T and in S, then it must satisfy the classical condition,

$$P(S|T) = \frac{P(S \wedge T)}{P(T)}$$

In other words, the odds you offer on a conditional bet must bear the appropriate relation to the odds you offer on a bet on the conjunction of the subject of the bet and its condition, and also on the condition alone. Of course, all the previous assumptions, provisos, and caveats apply. And this is hardly surprising, since we are talking of bets at a given point in time.

Conditional bets, so construed, are essentially just part of the static repertoire of betting. Everything is done at an instant of time. And just as straight bets can bear odds that conform to epistemological probability, and also to a coherent belief function, so also may (perforce) the odds on conditional bets conform to both epistemological probability and a coherent belief function.

This is mildly interesting, but not nearly as interesting as the claim that to be coherent (rational) you should adjust your betting odds in accordance with the principle of conditionalization applied *over time* as you get new information. Let us look more closely at this dynamic claim.

Suppose, hypothetically, that at time t you have a set of real-valued beliefs in a bunch of propositions that satisfies the probability calculus. That is, suppose that you are capable of posting odds on a field of propositions in such a way that the odds correspond to degrees of belief that satisfy the probability calculus. A new piece of evidence E is established between t and $t + \Delta t$. This leads you to change your distribution of beliefs among propositions in the algebra.

Assuming:

1. that E is among the elements in the algebra,

2. that to call E "evidence" entails that in our new state we assign probability 1 to E: $P(E) = 1$

3. that at $t + \Delta t$, our new beliefs should also be coherent,

what can we assume about the new distribution of belief? It turns out that without further assumptions, we can assume nothing at all! For example, suppose our assignment of beliefs to S and to $S \wedge T$ at t are .30 and .15, respectively. We now pass to $t+\Delta t$, and shift $P(T)$ to 1. Our new distribution of belief must assign $P(T) = 1.0$. But our assignment of belief to S remains totally undetermined. It could be .75.

To require that the new value of $P(S)$ be .5, i.e., to require that the new value of $P(S)$ be equal to the old value of $P(S|T)$, requires the invocation of a new principle baptised by Isaac Levi [8] "the Principle of Temporal Conditionalization," and "dynamic conditionalization" by Brian Skyrms [9]. Paul Teller [10] was the first to point out in print[3] that these principles needed a justification other than that provided by Ramsey for static conditionalization: that is, that something different was involved in making conditional bets at a *given* point in time, and in committing ourselves to new betting odds, conditional on the acquisition of new evidence during an *interval* of time.

We shall consider the new arguments for dynamic conditionalization shortly, but to begin with let us ask the question whether *any* change in belief between one time and another, during which certain evidence comes to be assigned probability 1, cannot be construed as the result of conditionalization.

The answer here is clearly "yes." Suppose our initial probability of H is $p(H)$, our initial probability for the evidence E is $p(E)$, and our initial probability for $H \wedge E$ is $p(H \wedge E)$. Clearly, after observing only E, we can take the new probability of H, $p'(H)$ to be different from the ratio of $p(E \wedge H)/p(E)$. One valid way of accounting for the difference between $p'(H)$ and $p(H \wedge E)/p(E)$ is to suppose our original assessment of $p(H \wedge E)$ and of $p(E)$ was reconsidered: on reflection we decided that these values *should have been* (before observing E) $p''(H)$ and $p''(H \wedge E)$, where of course $p''(H \wedge E)/p''(E) = p'(H)$. This remark is made even more poignant by the obvious fact that we cannot assign coherent probabilities to all statements at the drop of a hat; it requires some reflection, some computation, and even then we must be prepared to have made a mistake. We must be able to back up and reconsider.

Suppose we consider a hypothesis H, and some evidence E. In many

[3]Teller attributes the argument to David Lewis, who in turn credits Hilary Putnam.

cases—especially those involving statistical hypotheses and evidence—we can be quite confident of $p(E|H)$, the likelihood of H. The "prior" probability of H is another matter; this is notoriously subject to waffling and inconstancy. But since $p(H|E) = p(H)p(E|H)/p(E)$ we can make $p(H|E)$ vary from 0 to $p(E|H)/p(E)$ just by varying the prior probability we assign to H.

Furthermore, the likelihood $p(E|\neg H)$ is notoriously difficult to specify in a convincingly plausible way. By varying this, too, we can make

$$p(H|E) = \frac{p(H)p(E|H)}{[p(H)p(E|H) + p(\neg H)p(E|\neg H)]}$$

have any value we want between 0 and 1.

The upshot of these considerations is, first, that unless we begin with a definite coherent assignment of probabilities, it is difficult to know whether the updating after the observation of the evidence was purely due to the evidence (in which case conditionalization is intended to apply) or due also to further reflection on the degrees of belief entertained prior to the observation of E. Second, certain probabilities (the prior probability of H and the likelihood $p(E|\neg H)$ for example) are notoriously hard to determine. But lacking constraints on these probabilities, there are *no* constraints on the conditional probability of H given the observation E. We can always maintain that conditionalization is obeyed by taking the relation between the later probability and the earlier one to be telling us something about these hard to specify prior probabilities and likelihoods.

3 Dynamic Conditionalization

Teller [10] offers a dutch book argument which he claims supports dynamic conditionalization.As pointed out in the previous section, Ramsey showed that the odds posted on conditional bets must be derived from static conditional probabilities.That is, at any moment of time t_0 the agent's betting odds must satisfy the relationship

$$P_0(S|T) = P_0(S \wedge T)/P_0(T),$$

where P_0 is a distribution over the agent's beliefs at time t_0 that satisfies the probability calculus. The intent of Teller's argument is to show that the

agent's *new*, updated distribution of beliefs must be the result of conditionalization.That is, if at time t_1, E is the only piece of new evidence accepted by the agent, then the agent's new betting odds must satisfy

$$P_1(S) = P_0(S|E),$$

where P_1 is a new distribution over the agent's beliefs.

First, it is clear that all of the presuppositions of the static Dutch Book Principle presented in the previous section also apply here, i.e., real-valued beliefs, identification of degrees of belief with posted odds, willingness to take all bets, and fear of others taking advantage—along with their attendant difficulties. But, even if we grant these assumptions, it can be shown that Teller's argument for conditionalization as the rule of updating one's beliefs is not generally applicable (Teller does not claim that it is), and even in those special cases where it does apply it does not seem to offer much more than Ramsey's static argument in favor of conditional probabilities, that is, it seems to be more an argument for posting a certain set of static odds rather than a dynamic argument for a particular rule of update. To substantiate these claims let us examine Teller's dutch book construction in more detail.

There is an agent who has a belief function P_0 at time t_0. The belief function is assumed to be a real valued function defined over some set of propositions, and the agent is assumed to be committed to taking all bets at odds determined by this function.That is, for any proposition A in the domain of P_0, the agent is willing to buy or sell a bet which returns 1 if A and returns 0 otherwise for the price $P_0(A)$.

Another stipulation, which will turn out to be crucial, is that there is a set of events, $\{E_i\}_{i\in I}$,that are a mutually exclusive and jointly exhaustive set of propositions specifying in full detail all of the alternative courses of experience the agent might undergo between the time t_0 and the time t_1.

In the case that the agent knows at time t_0 what his belief function $P_1^{E_i}$ will be at time t_1 if E_i turns out to be true, and this new belief function is not the conditionalization of his old belief function, i.e., $P_1^{E_i}(A) \neq P_0(A|E_i)$, a bookie can buy and sell a set of bets (though not all at the same time) from our willing agent which will result in a net loss to the agent whatever happens. That is, the agent will be vulnerable to a dynamic dutch book. In addition to knowing all of the agent's betting odds, the bookie must also know something about the agent's new belief function $P_1^{E_i}$. In particular, the bookie must know if $P_1^{E_i}$ is less than or greater than $P_0(A|E_i)$.

Let's assume some particulars. Fix the particular E_i for which the agent's $P_1^{E_i}$ is not conditionalizing, and drop the superscript,i.e. identify P_1 with $P_1^{E_i}$. Let the proposition A (in the domain of the agent's original belief function P_0) be such that $P_1(A) \neq P_0(A|E_i)$, and let $P_1(A) < P_0(A|E_i)$.

With these particulars a dutch book is constructed in the following manner:Let $x = P_0(A|E_i)$ and $y = P_0(A|E_i) - P_1(A)$ (> 0).

At time t_0 the bookie sells the agent the bets

$$(a) = \begin{bmatrix} 1 & \text{if } A \wedge E_i \\ 0 & \text{otherwise} \end{bmatrix}$$

$$(b) = \begin{bmatrix} x & \text{if } \neg E_i \\ 0 & \text{otherwise} \end{bmatrix}$$

$$(c) = \begin{bmatrix} y & \text{if } E_i \\ 0 & \text{otherwise,} \end{bmatrix}$$

for the maximum price he will pay. This price is

$$P_0(A \wedge E_i) + xP_0(\neg E_i) + yP_0(E_i) = P_0(A|E_i) + yP_0(E_i).$$

Bets (a) and (b) together comprise a conditional bet on A; there is no gain or loss if E_i turns out to be false and it turns into a bet on A if E_i turns out to be true. The additional bet (c) insures that the agent will lose if the conditional bet is called off, i.e., if E_i is false the agent has a net loss of $yP_0(E_i)$.

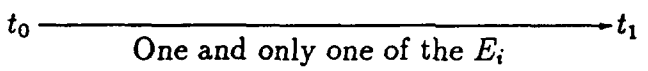
If E_i is true the agent now has an unconditional bet on A. At this time the bookie buys back this bet from the ever willing agent at a lower price. That is, he purchases a bet from the agent which pays 1 if A, 0 otherwise. Assuming that the agent determines his price for this bet from his new belief function P_1, the bookie can repurchase the bet at the price $P_1(A) = P_1(A|E_i) - y$. If the agent is willing to do all of this he will be left with a net loss of $yP_0(A|E_i)$ (positive) whatever happens; he has been dutched. If originally $P_1(A) > P_1(A|E_i)$ then the bookie follows the same procedure except he buys instead of selling and sells instead of buying.

After finding out that E_i is true the agent is left with a bet on A, he has paid out a total of $P_0(A|E_i) + yP_0(E_i)$, and has won a total of y (from bet (c)). If he chooses to sell his bet on A at the lower price determined by his new odds, he will take the certain loss of $yP_0(A|E_i)$. If he holds his bet then

he still has the prospect of winning 1 if A turns out to be true. In this case his winnings will total $1 + y$, a net gain. Why would the agent want to take the certain loss over the possibility of a net gain? A simple, and perfectly rational, reason is to buy himself out of a bad situation. The agent may well chose to minimize his new *expected* loss by taking a smaller certain loss now.

But does this mean that in acting rationally the non-conditionalizing agent can be dutched? No. A closer look at this betting scenario reveals that a logical agent would not be willing to enter into such a betting agreement in advance. It will be demonstrated that, again, it is a matter of deductive logic, not degrees of belief, that prevents the agent from being dutched.

To see this, let's examine some details of Teller's construction.Teller stipulates that there exists a partition of events $\{E_i\}_{i \in I}$ which occur between time t_0 and time t_1.



Now what about the original proposition A? Although Teller does not specify, it is clear that A cannot be one of the possible events in this time interval. Say it was, then there are two possibilities,$A = E_j, j \neq i$ or $A = E_i$ (where E_i is the event for which $P_1^{E_i}(A) \neq P_0(A|E_i)$). If $A = E_j$ then the agent will not pay any money for bet (a) and bets (b) and(c) will be zero payoff bets (in this case $P_1(A) = P_1^{E_i}(E_j) = 0$ since E_i and E_j are exclusive). No money will be won or lost. If $A = E_i$ then bets (a) and (b) will be unit bets for and against A, and bet (c) will be a zero payoff bet(in this case $P_1(A) = P_1^A(A) = 1$). In both cases no dutch will occur.

So we are left with the case that $A \notin \{E_i\}$. We have a situation in which at time t_0 the agent knows that A cannot occur until one of the $\{E_i\}$ occurs. The argument can now be construed as one concerning the odds initially to be posted on $A \wedge E_i$, irrespective of the agent's initial degrees of belief. It is a deductive constraint that the agent's posted odds satisfy

$$P_1(A) = P_1^{E_i}(A) = P_0(A \wedge E_i) \div P_0(E_i) = P_0(A|E_i).$$

But this may only mean that the odds posted on $A \wedge E_i$ can't correspond to the agent's initial doxastic state. That is,Teller's construction does not force

the agent to update by conditionalization any more than it forces the agent to post certain static odds.

In summary, the construction only applies to a very special situation, in which the agent can used his deductive abilities to avoid being dutched.

4 Justifying Conditionalization

Teller [10] seeks to provide also a qualitative argument for conditionalization deriving from conditions on reasonable changes of belief. He characterizes the problem of justifying conditionalization as the problem of supporting the claim that under certain well-specified conditions, ***only*** changes of belief described by conditionalization on some proposition which an agent has come to believe true (for example, by means of observation) are ***reasonable*** changes of belief; that no change of belief that is not described by conditionalization on some proposition is a reasonable one. Teller's argument takes the following form. First, he shows that the conditionalization criterion, which he calls '$Cond(E)$', the requirement according to which change of belief takes place by conditionalization on proposition E, more exactly

$$Cond(E) =_{df} 0 < P_0(E) < 1 \wedge (A)[P_1(A) = P_0(A|E)],$$

where P_0 is the agent's old belief function and P_1 is his new, updated belief function, is equivalent (given certain strong assumptions about the structure of the body of belief) to $C(E)$, where

$$\begin{aligned} C(E) \;\; =_{df} \;\; & 0 < P_0(E) < 1 \wedge P_1(E) = 1 \wedge \\ & (A)(B)[\text{if } ((A \vdash E) \wedge (B \vdash E) \wedge P_0(A) = P_0(B)), \\ & \qquad \text{then } P_1(A) = P_1(B)] \end{aligned}$$

in which '$\not\vdash$' is used to mean 'logically implies.' Let us grant Teller this equivalence for the sake of argument.[4] Now $C(E)$ is the constraint that the agent's belief in E changes from something greater than zero to unity, and furthermore, for any two propositions, A and B, each of which logically implies E, if A and B are believed to the same degree by the agent before E

[4] We won't quibble with the assumptions required by Teller to show that $Cond(E)$ is equivalent to $C(E)$, but point out that these assumptions are by no means intuitively obvious.

is learned, then they must be believed to the same degree upon learning E. Now Teller defines 'reasonable change of belief' as follows:

(D1) A change of belief is reasonable if and only if

- **(a)** the new beliefs are reasonable, or
- **(b)** both the new and old beliefs are not reasonable, but the new beliefs would have been reasonable if (i) the old beliefs had been reasonable and (ii) both before and after the change the agent has a high reasonable degree of belief that his old degrees of belief were reasonable.

Teller then proceeds to argue for conditionalization in the following way. First, he asserts a principle **P**.

P Let E be any proposition such that

- **(a)** The agent's initial degrees of belief are reasonable.
- **(b)** Initially the agent is unsure of the truth of E.
- **(c)** The agent comes to know that E is true.
- **(d)** After coming to know that E is true, any reasons the agent might have which in fact make reasonable or justify changes in other beliefs are either directly given by or included in his new knowledge that E is true; or such reasons indirectly rest on his new knowledge that E is true.[5]

Then for any two propositions A and B, such that

- **(e)** A and B each logically imply E
- **(f)** The agent's initial degree of belief in A and in B are the same

[5] In Teller's idiom, a belief whose object is proposition X is ***directly given by*** a belief whose object is Y just in case $X = Y$ or $X = \neg Y$; a belief whose object is a proposition X is ***included in*** the belief whose object is proposition Y if X is a conjunct of Y or X is the negation of a conjunct of Y; a belief whose object is proposition X ***indirectly rests*** on a belief whose object is proposition Y just in case the agent has arrived at his belief that X is true by a chain of reasoning whose initial premises are all directly given by or included in his belief that Y is true.

it is also the case that

(g) The agent's new degrees of belief in A and in B are reasonable only if after coming to know that E is true they continue to be the same.

Teller then claims that if **P** is true, and if conditions (a)-(d) hold true of a proposition, then the new beliefs are all reasonable only if $C(E)$. Now we have granted that $C(E)$ if and only if $Cond(E)$. So by (D1), if the initial beliefs are reasonable, change of belief is reasonable only if the new beliefs are reasonable. So if principle **P** is true, and if conditions (a)-(d) hold of a proposition E, then change of belief is reasonable only if it is described by conditionalization on E.

The argument is a valid one. So Teller's justification of conditionalization goes by way of justifying principle **P**. Teller claims that **P** is plausible and interesting, and worthy of critical scrutiny in its own right. Furthermore, he proposes to defend it by example. But he never claims that it is true. We claim that it is false in some relevant cases, and that it does not even apply to cases where it is needed to justify conditionalization. We propose to show that this is so by turning the very example Teller offers up in its favor *against* principle **P**.

First, however, we suggest that there is a simple case of an agent who violates **P** but who is nevertheless rational, from an intuitive point of view. Consider an agent whose assignments of belief include the following:

Scan $P(Peterson\ is\ Scandinavian) = 1$

Swede $P(Peterson\ is\ a\ Swede) = 0.2$

Nor $P(Peterson\ is\ a\ Norwegian) = 0.8$

Stat $P(80\%\ of\ all\ Scandinavians\ are\ Swedes) = .2$

That is, initially the agent knows that Peterson is Scandinavian, believes it likely that he is Norwegian, unlikely that he is Swedish, and that it is unlikely that 80% of all Scandinavians are Swedes. Let us also assume that Swedes and Norwegians are the only kinds of Scandinavians, they are mutually exclusive, and furthermore the agent knows this.

Let

$$\begin{aligned} A &= Swede \wedge Stat \\ B &= Nor \wedge Stat \\ E &= Stat \end{aligned}$$

Say the agent initially holds an equal degree of belief in the two assertions A and B, and he can certainly do so while still having beliefs which satisfy the axioms of probability (for example, the agent could have other evidence that these conjunctions are equally likely). Let the degree of belief in A and B be, say, 0.1. Upon learning that E is in fact true (i.e., that 80% of all Scandinavians are Swedes) each of A and B becomes equivalent to the simpler assertions that Peterson is a Swede and that Peterson is a Norwegian, respectively. In light of this new statistical information it seems quite plausible that the agent may wish to believe A more than he believes B. Clearly both A and B imply E; hence this situation is a violation of principle P. But there is nothing clearly objectionable about the rationality of our agent.

The Bayesian might argue that the the same statistical facts that make the *new* degree of belief in A larger than that in B are asserted as part of A and B, and should initially incline the agent to believe A more strongly than B. But there may be other considerations—e.g., that Peterson is so typical of all those whom the agent takes to be Swedes as to seriously undermine the hypothetical statistical statement. In any even we will turn aside from this argument to the argument promised earlier: we will turn Teller's own example against **P**.

Teller's example is as follows: Suppose that two men are going to race and a given agent has equal degrees of belief in two propositions, A and B; A reports that the first man wins, and B reports that the second man wins. The agent cannot be certain of E, the proposition that one of the two men wins, because he recognizes that the race might be called off or might result in a tie. Now the agent learns, and learns no more than, that E is the case: that the race is successfully completed and does not result in a tie. "Under these conditions," claims Teller, "it would be absurd for him now to shift his beliefs so that he is rather more confident in A than in B or in B than A."

Not so absurd, we claim. For suppose that our agent is acquainted with the two racing men in question. He knows that the first man, 'a' let's call him, is a self-absorbed, over-confident sort; and that the second man, 'b'

for short, needs very badly for personal reasons to win this very race (say, for instance, in order to save his dying daughter's life). Our agent knows that a and b have comparable track records and physical abilities, and that they have trained equally rigorously for the race in question. Furthermore, our agent is inclined to disregard the psychological states of rivals as a poor and unreliable indicator of outcomes because his evidence in connection with the weight of psychological factors is mixed. And he regards his assessment as reasonable. Furthermore, it is reasonable. So our agent is exceedingly confident of his belief that the race, if run, would result in a tie. He believes propositions A and B to the same degree, namely zero (or, if you prefer, let him be confident in A and B equally to degree 0.0017; the argument is equally effective in either case). Each of A and B logically implies E, the proposition that some man wins the race; whether our agent believes either one of A and B has no bearing at all on the question whether each of A and B logically implies E—they do the implying independently of whether our agent has any confidence in either one of them.

Now the agent learns that the race is successfully completed, that it has taken place and has resulted in a victory for one of the participants in question; and that is all that the agent learns. Now our agent pursues the following train of thought.

> Let me suppose that all the evidence I have is all that is *relevant* to the outcome of the race (that is, that the race did not result in a victory simply because of something unexpected, for instance that one of the participants was struck by a fatal heart attack in mid-race). My prior assessment of the evidence in my possession indicated that the race should have resulted in a tie; and I believed as much. But it did not. Hence my prior assessment of the probabilities at issue, though it was not unreasonable then and continues to be reasonable now, must have been slightly mistaken. And if I'm right that I am in possession of all the facts relevant to the outcome of the race, then my assessment is mistaken because I failed to give proper apportionment of weight to all the evidence in my possession. I must have been mistaken to discount the psychological states of the participants. As I recall, participant b had overriding additional motivations to win the race which participant a did not. This must have figured more

> largely in the outcome of the race than I had previously thought it would. Hence I hereby change my mind concerning the weight of psychological factors of participants in races. Henceforth I will regard proposition B to be more likely true than proposition A.

The foregoing, we suggest, is not only lacking in absurdity, but it is also eminently reasonable. And if we are right about this, then it is both the case that Teller's principle **P** is false and that conditionalization upon evidence is by no means the only reasonable way to change belief.

It could be claimed that **P** did not apply: that the agent employed or arrived at a new piece of data after having learned E, which datum does not logically rest, directly or indirectly, on the new observation E, as **P** ordains. Now to resurrect **P** as the truth in this case, one needs to show that there is a way of describing the case from a philosophical point of view according to which **P** is not contravened. This redescription involves claiming that what the agent comes to learn is not merely E, but also the datum cited by the agent in arriving at the conclusion he does, namely, the datum, call if 'V', which mandates that old evidence, already in the agent's possession be reassessed and rehabilitated for purposes of conditionalization. By means of V, then, old evidence comes to be treated as "new" evidence and becomes grist for the conditionalizer's mill. The hard part of the Bayesian's case is to show that to invoke V is really to employ new evidence. By all reasonable standards, V is not any kind of evidence at all. The observation is E; E is the evidence. And it is clear that the agenet never "learned" V at all; he may have had "reassessment" principles all along as part of his "operating system" or "software." It is just that learning E in this situation propted him to invoke V at the point he did. The Bayesian needs strong argument to establish that V really is evidence of some kind or other. In the absence of such and argument the judgment must be that a defense of this stripe conflates evidence with the mechanisms by way of which evidence is evaluated: the Bayesian who says that **P** applies in the way we have outlined confuses evidence with the import of evidence. This makes it look as if evidence cannot be identified until a change of belief is effected. And it makes it look like the same observation can constitute different "evidence" for different agents—a very unsavory and counterintuitive result by all standards. And hence the rebuttal fails.

But if the rebuttal fails, then the objection stands: **P** does not apply to our case in which an agent fails to update by means of conditionalization

and must nevertheless be judged rational. But if this is right, a defense of **P** does not constitute a justification of conditionalization, even if **P** is true.

What do we learn from the forgoing example of the "reflective" agent? That it is reasonable to respond to evidence by changing one's views concerning the impact of previous evidence. And that this kind of changing of one's views is not a kind of conditionalization.

This observation can throw light on two aspects of scientific inquiry that have been recently noted. First, it makes sense of how anomalies in scientific theories are tolerated for periods of time; the explanation is that what we *now* see as anomalous was *then* (reasonably) taken to have little evidential import. Second, it makes sense of the way in which relatively small pieces of evidence, that were available all the time, are suddenly considered fatal to an older theory; the explanation here is that new evidence bears on the significance of the old evidence: the import of the old evidence has changed in the light of a new piece of evidence.

So far we have argued that conditionalization is by no means the *only* rational course for belief change. Now we will argue that under certain conditions, it is *never* a rational course of action to change belief by conditionalization. The certain conditions we have in mind are those in which rational agents are are in irrational states of belief (where 'irrational' does not indicate being in a state of belief not described by the probability axioms). We will argue that to change one's beliefs always by conditionalization on evidence is to determine once and for all the impact or import of evidence. And for the temporarily irrational believer, this is epistemically fatal.

If a believer starts out doxastic life with an unreasonable set of beliefs, there is no telling when, if ever, that believer may achieve rationality just by conditionalizing on new evidence. Consider an agent who believes an outright contradiction, and suppose that this agent is a perfect logician. If this believer is in possession of contradictory beliefs, then she will know this fact about herself. Now if conditionalization is the truth about rational change of belief, then such a believer has no rational way of simply "converting" to rationality. So in the case of this believer, we are inclined to say that conditionalization is *never* a rational way to change her belief. The exceedingly rational option, and the *only* rational one available to her in our view, is just "conversion" to rationality.

What's that you say, gentle reader? You think that it is just not possible for someone to believe a contradiction? All right. But surely you believe

that it is possible that someone be in possession of a distribution, call it P, over an algebra of beliefs which, though it does not yield a contradiction outright, is nonetheless incoherent—in the technical sense that it violates the probability axioms. Now this unfortunate believer can never come to have coherent beliefs merely by conditionalization. How is this so? Let P' be any member of the set of probability distributions over the set of sentences in our poor believer's body of belief which are coherent. But since P' is coherent and P is not, it can *never* be that

$$(\star) \qquad P'(A) = P(A|\bigwedge E_i) = P(A \wedge \bigwedge E_i)/P(\bigwedge E_i),$$

where '$\bigwedge E_i$' names the set of all those propositions which our unhappy agent ever does (or can, if you like) come to learn and upon which she conditionalizes; for P is just incoherent, by hypothesis, and if $(\star)$ were true, then our hypothesis would be false and the example altered. Hence the incoherent conditionalizer can never achieve coherence.

Now we should think that if one advocated coherence (in the sense that one championed the probability axioms in one's own doxastic life and enjoined them upon others), then one would and ought to say concerning the incoherent believer who knows himself to be incoherent, that in his case it is *never* a rational change of belief to change belief by conditionalization. We do not tout the probability axioms in the same way, but even so we say this: the *only* rational course of action for a believer who believes irrationally and knows himself to believe irrationally is to "convert" to rationality.

5 On (Reflection)

Bas van Fraassen considers the case of a certain fictitious mortal, let us call him 'Dupe,' whose story he tells in the first person for drama's sake [11]. Dupe does not today profess faith in the basic theory of evolution, nor is he certain whether he shall do so one year from today. But it seems to him eminently possible that he shall come to believe as the Darwinians believe; and, moreover, it also seems possible to him that he will come to believe in Darwin's theory while Darwin's theory is false.

Van Fraassen shows us that these beliefs of our Dupe are, in fact, incoherent—today. He does this by demonstrating to us that, no matter how Dupe arrives at the beliefs which he will hold one year from today, a Dutch bookie can

offer him a set of wagers which he today considers fair, according to which he loses no matter what happens a year from today. The strategy which van Fraassen suggests that the bookie may use is as follows. Let H be the thesis that Darwin's theory is true and E be the proposition that Dupe will come to believe H one year from today. Let Dupe's degree of belief in E, $P(E)$ for short, be equal to 0.4, and his belief that he will mistakenly come to believe H, $P(\neg H \wedge E)$, be equal to 0.2. (Grant also that full belief corresponds to personal probability equal to 1.) Now the bookie offers Dupe the following set of three wagers: the first pays 1 if Dupe comes to believe H and H is really false—and asks 0.2 for it. The second pays 0.5 if Dupe does not come to believe H, and he asks 0.3 for that. The last pays 0.5 if Dupe should come to believe H, and costs 0.2. None of the bets pays anything if not won; each is considered fair on Dupe's reckoning, and their total cost is 0.7. More generally, the three wagers can be characterized as follows:

(I) The bet which pays 1 if $(\neg H \wedge E)$ and which costs $P(\neg H \wedge E)$;

(II) The bet which pays x if $\neg E$ and which costs $xP(\neg E)$;

(III) The bet which pays y if E and which costs $yP(E)$;

where x is the conditional probability of $\neg H$ on E, and y is x minus the subjective probability the customer will have for H, when and if E becomes the case. I and II together form a conditional bet on $\neg H$ given E, a bet costing x and bearing prize 1, a bet called off if E turns out not to be the case. The total cost of all three bets is $x + yP(E)$.)

Now the possible outcomes are two. On one scenario Dupe does not come to embrace H; he wins the second bet and loses the other two. On the other he does come to embrace H; he loses the second bet, reports himself that H is true, and so is willing at that time to sell a bet back to the bookie on $\neg H$ for next to nothing, and wins the third bet. In either case Dupe receives a little more than 0.5, and has netted a loss. The bookie has devised a strategy based upon knowledge available to Dupe himself, a strategy allowing the bookie to offer Dupe only bets that Dupe presently considers fair and yet would necessarily result in certain loss for him. Dupe's vulnerability to Dutch book, concludes van Fraassen, makes Dupe's initial state of opinion a demonstrably bad guide to life.

But if Dupe is diachronically incoherent, then so is every experimental scientist who admits to fallibility. To see that this is so, let e be the sort

of proposition that a scientist typically takes to report evidence; perhaps it is 'The current is 2.47 + 0.02 amps' or 'The solution turns a pale shade of pink'. Now if our scientist is reasonable, she will confess that she is not infallible when it comes to perception (and observation more generally) and so may be mistaken when she accepts e. Her subjective probability that e is false on the supposition that she takes it as evidence is not zero. But then it looks as if she becomes vulnerable to being Dutched in the same way Dupe is. Let E be the proposition that our scientist takes e as evidence, and H the thesis that e is true. And let our scientist assign definite values to $P(E)$ and $P(\neg H \wedge E)$ in accord with her suspicion that she is not infallible.

Van Fraassen's way out of this difficulty is to declare both Dupe and our scientist as genuinely irrational because in each case our protagonist's degree of belief about what would happen, on the supposition that he or she would have certain opinions concerning the events in question, differs from the present opinion concerning those events. That is, our protagonists fall afoul of (Reflection):

$$\text{(Reflection)} \qquad p_t(A|p_{t+x}(A) = r) = r.$$

Here p_t is the agent's credence function at time t, x is a non-negative number, and $(p_{t+x}(A) = r)$ the proposition that at time $t+x$ the agent bestows degree of credence r on the proposition A. To satisfy (Reflection), the agent's present subjective probability for proposition A, on the supposition that the agent's degree of credence in A at some later time is r, must equal the number r.

van Fraassen claims that satisfaction of (Reflection) ensures satisfaction of the probability calculus, in the sense that it necessitates invulnerability to Dutch book. We have already argued that Dutch book arguments are deeply suspect as indicators of the truth about what we ought to believe. The fact that both (Reflection) and Dutch book strictures result in degrees of credence which obey the probability axioms does not show that Dutch book or (Reflection) reveal the truth about how we ought to *believe*. Our doxastic heros need not be bullied or browbeaten into *believing* in accordance with probability axioms by means of Dutch book or (Reflection); they, being perfectly deductively endowed, simply will refuse to accept a set of wagers that will result in certain loss for them.

But, it might be argued, one ought to believe in such a way that one gives oneself the chance that one's beliefs are "correct"—in the sense of being perfectly calibrated with actual frequencies. (Reflection), argues van

Fraassen, requires the agent to express perfect confidence in the calibration of her judgment. Perfect calibration, according to van Fraassen, is perfect agreement between an agent's judgment and actual frequencies; a forecaster, for example, is perfectly calibrated if, for every number r, the proportion of rainy days among those days in which he announces rain with probability r, is just r.[6] Now this criterion of perfect calibration can be shown to be exactly equivalent, claims van Fraasen, to satisfaction of the probability calculus (in the same sense that this equivalence can be claimed for the criterion of invulnerability to Dutch book). And "it would seem to be irrational," van Fraassen suggests, "to organize your degrees of belief in such a way as to ruin, a priori, the possibility of being perfectly calibrated."

But what exactly is so irrational about arranging one's doxastic life so as to insure the failure of perfect calibration? van Fraasen himself admits that calibration may not by itself be a reasonable aim. And it is not unreasonable to suppose that if one managed to arrange one's doxastic affairs so as to give oneself the chance to be perfectly calibrated (even on van Fraassen's notion of calibration), it can only have happened so purely by accident.

We believe that (Reflection) is unreasonable. W.J. Talbott has presented several convincing counterexamples to it [13].[7] The following is one of Talbott's examples. Suppose that I plan to attend a party tonight, and I plan, furthermore, to become inebriated. Let A name the proposition which reports that at $t + x$ (that is, at the end of my evening), my reactions and driving performance are seriously impaired. Presently (that is, at t) I believe A to a degree of confidence on the order of certainty. I presently also believe with just as great a degree of confidence that at $t + x$ (when I am sufficiently tight at the end of the evening) I will not believe for a moment that my reactions are impaired. So I am described by the following pair of propositions:

(i) $p_t(A) = 1$

(ii) $p_t(p_{t+x}(A) = 0) = 1$.

[6]Seidenfeld [12] has shown that this *cannot* be an adequate notion of calibration. One could predict rain with probability p every day, where p is just the proportion of rainy days so far that year.

[7]The authors did not have access to anything more an abstract of Talbott's paper. We have drawn our own conclusions from his examples; conclusions that may or may not be similar to the conclusions arrived at by Talbott.

Now applying (Reflection), I am caught in contradiction, for by (Reflection) I am required to be described by

(iii) $p_t(A|p_{t+x}(A) = 0) = 0.$

And so by (ii) and (iii):

(iv) $p_t(A) = 0.$

Now (i) and (iv) together constitute a contradiction. van Fraassen urges that we *accept* (iv) and therefore reject (i). That is, he insists that I ought not at all to believe *now* that my driving performance at the end of the evening will be impaired (despite all my resolve to get smashed). But if *anything* is a demonstrably bad guide to life, this is. The rational agent ought *not*, we claim, be described by (iv) but rather by (i). And if we ought reject (iv), then we ought reject the faulty principle which makes it derivable, namely (Reflection).

But van Fraassen is not without recourse at this juncture. It is open to him at this point to argue that (Reflection) applies only to *reasonable* beliefs. The belief at $t+x$, when I am entirely too snookered to see anything straight, ought not to be considered by me at t to constitute *any* kind of reasonable evidence as I reflect upon myself at t. And so (Reflection) does not deliver up any injunctions whatever with respect to the proposition A in question. (iii), in particular, is not an ordinance of (Reflection); so (iv) is not derivable by (Reflection), and it is perfectly reasonable for me to believe (i).

We are not entirely persuaded by such a rejoinder. It is certainly the case that I will at the end of the evening believe myself completely competent to drive; this is just one more fact, as good as any other, about me. Why should it not count from (Reflection)'s point of view? Why is it not respectable as evidence about my future self? Presumably the answer is that the belief is unreasonable at $t + x$. But what exactly is so unreasonable about believing, when one's wits are completely numbed by alcohol, that one is in full possession of one's faculties, particularly when that belief is thrust upon one from without? To say that a person's beliefs at a given time are unreasonable is to say that in some sense the person in question ought not to believe as she does at that time. But if ought implies can, then if it is the case that one is not able due to drink to believe otherwise than one does, then it cannot

be anything but false to say that she ought nevertheless to believe otherwise. There is a sense, therefore, in which a drunken person's beliefs are not unreasonable if they are due to the effects of drink.

But even if this response to predictable drunkenness is considered adequate, there is another case, again due to Talbott, which shows (Reflection) to be questionable: a case where unreasonableness is surely not the explanation. In this case we are to suppose that today I have spaghetti for dinner. Let the proposition S report this fact. Let t be today and x be a span of one year. And let '$\bigwedge E_i$' name the set of propositions I will come to learn between today and one year hence. Suppose furthermore that I generally eat spaghetti for dinner one out of ten times, and that I am aware of this fact about myself. Now I am quite confident today that today I have spaghetti for dinner. And surely anything that I will come to believe in the next year can have no bearing on the confidence with which I believe S today. So I am described by the following set of sentences:

(1) $P_t(S) = 0.99$;

(2) $P_t(S|\bigwedge Ei) = P_t(S)$.

Now it is reasonable to think that I shall forget one year from today what I have for dinner today; that the evidence which is clearly in my purview today shall not be so clearly etched in my memory one year hence. Moreover, it is reasonable to be forgetful. In addition, it is the case that human beings cannot help but be forgetful in this way; they are, of a piece, describable as forgetful in this sense. Suppose that I believe this about myself today. Then

(3) $P_{t+x}(S) = .1$,

(4) $P_t(P_{t+x}(S) = .1) = .99$.

(3) reports that I will forget that I had spaghetti one year hence and so the only evidence I will have to go on concerning what I had for dinner today is the frequency with which I eat spaghetti generally, and (4) reports that today I am practically certain of this fact.

But (Reflection) would have it otherwise. It demands:

(5) $P_t(S|P_{t+x}(S) = .1) = .1$.

Hence by (3) and (5)

(6) $P_t(S) = .1$.

But (6) contradicts (1). And surely it is much more reasonable that I be described by (1) rather than by (6): I am as certain as I can be today that this day I had spaghetti for dinner! (Reflection), in conjunction with my other beliefs about myself, (3) in particular, entails (6). So either (3) is unreasonable or else (Reflection) is. We suggest that the culprit is (Reflection). For surely (3) is nothing but good self-knowledge; it reflects an accurate assessment of my ability to recall the contents of my repasts long gone by.

The only recourse a defender of (Reflection) has available in connection with this assault is to respond that (Reflection) assumes an agent with perfect recall—an agent for whom once a piece of evidence is available to her at t, it is forever after available to her in the same way. But at this stage we begin to wonder whether the defenders of (Reflection) can have in mind anything like a human being as agent. First they claim that I am not reasonable when, drunk, I believe I am capable of driving, though I cannot help but believe as I do. Then they require of me that every piece of evidence I ever acquire be forever after available to me as on the day I first acquire it, though, again, I cannot possibly meet with this demand. Why not go one more, defenders of (Reflection)? Why not simply demand of me that I believe all truths and only truths, and consider me unreasonable otherwise? I am just as able to meet this demand as I am to meet the others.

"Idealization," it will be argued, is inevitable. Of course. But in order to take an idealized normative theory seriously, we need to be shown how to take steps toward the idea. "Forget less," "Don't get drunk," and "take account now of all the evidence that you ever could think of as relevant." are not useful injunctions. This is not to say, nor do we mean to have the reader think, that temporal conditionalization is always misleading or irrational. There is one set of circumstances—also idealized—under which even frequentists think conditionalization is correct. Namely, when we are dealing with a joint distribution of two quantities, and we happen to know the value of one of them. This is less general, and still idealized. And how to connect frequencies to beliefs is a tricky question (not solved by reflection or anything like that).

Why do we claim that (Reflection) is false (with respect to human beings)? For many the same reasons that we reject diachronic conditionaliza-

tion as the correct rational procedure for change of belief (again in humans). And it is this: that it is simply epistemically wrong to determine once for all time, a priori, the *impact* of evidence. For instance, in the first of Talbott's examples (the drunken case), I ought not at t to give any weight at all to the evidence reported by the proposition that at the end of the evening I will disbelieve that I am impaired. I will have learned from *experience* not to give weight to such evidence as A reports. In the second case (the spaghetti case), it is not unreasonable for me to believe that I will forget that today I eat spaghetti, that the evidence on the basis of which I believe S today will not be as clearly etched in my memory as it is today. Furthermore, how can it be unreasonable for me to forget evidence when I cannot help but forget? And if I cannot help but forget, then it is not unreasonable that one year from today I shall give little if any weight to the evidence upon which I today believe S; that evidence simply will not be available to me. (Reflection) treats the date of evidence as irrelevant. Hence it gets the wrong answers concerning what a rational agent ought to believe.

6 Carnap's Suggestion

An "inductive method" for Carnap is characterized in part by a constant λ, roughly giving the relative importance of empirical and logical factors in statistical induction. Lambda may range from 0 to ∞, where the latter value implies that observed relative frequencies have no influence on our probabilities, and the former implies that only relative frequencies, and not relative width, have influence. In Carnap [14] the lambda functions provided a unique characterization of inductive methods—i.e., of prior probabilities, subject to the relatively basic constraints of Carnap's system.

Carnap writes, "We regard an inductive method characterized by λ as the more successful in k [a specific state description], the smaller the mean square error of the estimates supplied ... for the relative frequency of the strongest properties expressible in [our language]." [p.2] Thus "Questions concerning the success of a given inductive method in the actual world would be of a factual, nonlogical nature." [p. 59]

In the case of the singular prediction that the next entity has the molecular predicate M, given that the relative width of M is w/k and that of a sample of size s, s_M have had the property M, the degree of confirmation is

given by:

$$\frac{[s_M + (w/k)\lambda(k)]}{[s + \lambda(k)]}$$

Carnap suggests, in the appendix of (1952), that a value of λ well worth considering is $\lambda = (s)^{\frac{1}{2}}$. Of course, to change λ is exactly to shift prior probabilities, and thus to violate the principle of temporal conditionalization, and thus to be in a position to have a dutch book made against one. If you adopt this principle, it is clear that a dutch book can be made against you as follows:

If, as background, we have observed 4 objects, of which 2 have had a property P of width 2, where the number of Q-predicates is 8, our probability that the next object will have the same property, $P(5)$, is

$$p_0(P(5)) = \frac{[2 + (1/4)\lambda(k)]}{[4 + \lambda(k)]} = \frac{2.5}{6} = .417.$$

Our probability that the next *two* will have P is the probability that the next will have P multiplied by the ***conditional probability*** (now, while we still have only a sample of 4) that the sixth item will have P given that the fifth does. i.e., multiplied by

$$p_0(P(6)|P(5)) = \frac{[3 + 1/4\lambda(k)]}{[5 + \lambda(k)]} = \frac{(3.5)}{7} = .500.$$

This, for $\lambda(k) = \sqrt{3} = 2$ is $.417 \times 3.5/7 = .208.$, i.e., $p_0(P(5) \wedge P(6)) = .208$.

But *after* we have observed that the fifth item has P, the probability that the sixth has P is

$$p_1(P(6)) = \frac{[3 + (1/4)(\sqrt{5})]}{[5 + \sqrt{5}]} = \frac{3.559}{7.234} = .492.$$

We may now follow the recipe provided by Paul Teller and spelled out above to explicitly dutch Professor Carnap:

1. Sell a bet on $P(5) \wedge P(6)$ that returns a dollar for the fair price of .208.
2. Sell a bet on $\neg P(5)$ for the fair price of .583.
3. If $P(5)$ occurs, buy back the bet on $P(5) \wedge P(6)$ (which is now simply a bet on $P(6)$) for .492—an amount less than Carnap paid for it.

Have we found a giant nodding? Has Rudolf Carnap, of all people, been found wanting in rationality? Surely not. The simple solution is that a set of bets must be made on the basis of a fixed value of λ. Carnap is not committing himself, and need not commit himself, now, to posting odds in the future that are conditionalizations of odds that he is willing to post now. He leaves open to himself the possibility of changing λ, and even changing λ in a predictable way. That he will not be dutch booked goes without saying—this is a deductive matter, independent of both his degrees of belief and how he changes them. That he allows his inductive logic itself to be influenced by the course of his experience requires that he sometimes changes his beliefs in ways other than by conditionalization. This need not reflect inductive inconsistency, but only sensible flexibility.[8]

7 Conclusion

Teller's **P** and van Fraassen's (Reflection) both constrain an agent to fail to be teachable concerning the weight of evidence. So if **P** and (Reflection) are right, they are not right about *human* agents. If they are right, they are not relevant to the truth about human epistemology. Humans are changeable by necessity; they are teachable because they are changeable. Hence one proper subject of human epistemology insofar as it deals with change of belief is precisely what is the right way for humans to bew teachable with respect to the import of evidence—just what is ruled out by temporal conditionalization and (Reflection).

Have we, by lifting the strictures of conditionalization and Reflection from our human believers, thereby consigned them to the mercy of dynamic Dutch bookies? Clearly not. As we have argued from the beginning, that an agent ought not to accept a set of wagers according to which she loses come what may, if she would prefer not to lose, is a matter of deductive logic and not of propriety of belief. Even a mechanical Carnapian procedure for changing probability functions need not be irrational.

[8]It should be pointed out that the proposal in question was made in (1952). In the more recent studies, [15] and [16], it is not reiterated. Carnap had moved more toward the personalist camp. But it is not repudiated, either.

References

[1] Henry E. Kyburg, Jr. Subjective probability: Considerations, reflections, and problems. *Journal of Philosophical Logic*, 7:157–180, 1978.

[2] Charles Chihara and Ralph Kennedy. The dutch book argument: Its logical flaws, its subjective sources. *Philosophica Studies*, 36:19–33, 1979.

[3] Patricia Baillie. Confirmation and the dutch book argument. *British Journal for the Philosophy of Science*, 24:393–397, 1973.

[4] F. Schick. Dutch bookies and money pumps. *Journal of Philosophy*, 83:112–118, 1986.

[5] F. J. Anscombe and R. J. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34:199–205, 1963.

[6] Henry E. Kyburg, Jr. *The Logical Foundations of Statistical Inference.* D. Reidel, 1974.

[7] F. P. Ramsey. *The Foundations of Mathematics and Other Essays.* Humanities Press, New York, 1931.

[8] Isaac Levi. Confirmational conditionalization. *Journal of Philosophy*, 75:730–737, 1978.

[9] Brian Skyrms. Dynamic coherence. In MacNeill and Umphrey, editors, *Foundations of Statistical Inference*, pages 233–243. Reidel, Dordrecht, 1987.

[10] Paul Teller. Conditionalization and observation. *Synthese*, 26:218–258, 1973.

[11] B. Van Fraassen. Belief and will. *Journal of Philosophy*, pages 235–256, 1984.

[12] T. Seidenfeld. Calibration, coherence, and scoring rules. *Philosophy of Science*, 52:274–294, 1985.

[13] W. J. Talbott. Reflections of two principles of bayesian epistemology. In *APA Eastern Division Colloquium on Logic, Probability and Methodology*, 1987.

[14] Rudolf Carnap. *The Continuum of Inductive Methods.* University of Chicago Press, Chicago, 1952.

[15] Rudolf Carnap. A basic system of inductive logic part I. In R. Carnap and R. C. Jeffrey, editors, *Studies in Inductive Logic and Probability I*, pages 33–165. 1971.

[16] Rudolf Carnap. A basic system of inductive logic part II. In R. C. Jeffrey, editor, *Studies in Inductive Logic and Probability II*, pages 7–155. 1980.