

THE DAY OF THE DOLPHINS

PUZZLING OVER EPISTEMIC PARTNERSHIP¹

Bas C. van Fraassen

Published pp. 111- 133 in A. Irvine and K. Peacock, eds., *Mistakes of Reason: Essays in Honour of John Woods*. University of Toronto Press, 2005.

| | |
|---|----|
| I. Background: a position in the philosophy of science..... | 2 |
| II. Observability perspectival..... | 3 |
| Observation and the epistemic community | 3 |
| Smart detectors and bionic persons..... | 4 |
| Changes in our epistemic community | 7 |
| The new riddle of prevision | 9 |
| Epistemic marriage | 10 |
| III. Self-Prevision: its logical laws, its subjective sources | 12 |
| Hintikka's problem | 12 |
| Logic in the first person | 13 |
| Rational change in view..... | 17 |
| IV. Epistemic marriage revisited | 22 |
| The Reflection Principle applied | 24 |
| To envaguen..... | 24 |
| Conclusion | 25 |
| NOTES..... | 27 |

It is a curious but profoundly important fact that general philosophical problems, no matter how traditional or venerable, lead us into sticky technical problems. In fact, the topic of this paper is a technical problem about subjective probability reasoning, but I got into it quite innocently by taking a position in philosophy of science. It was an unpopular position so I had a lot to defend. Today – well, today after many years of struggle to vanquish the foes, and much

The Day of the Dolphins

son et lumière on both sides, ... it is still unpopular and there are still more technical problems ... but hope springs eternal

I will begin with a brief introduction to the general philosophical problems and then jump into subjective probability reasoning about what our future can be if we think we may have surprisingly alien new partners in the enterprise of knowledge.

I. Background: a position in the philosophy of science

The position I took is that full acceptance of science, even with no qualifications and no holds barred, need not involve belief that the sciences are true – that even such wholehearted acceptance requires no more belief than that they are empirically adequate. That means: what the sciences say about the observable parts of the world is true, the rest need not matter. I'm putting this very roughly, but it is enough to make you see the immediate challenge. Suppose that in accepting science I believe whatever it says about the observable. Doesn't the line between the observable and the rest of nature shift continuously – with the invention of microscopes, spectroscopes, radio telescopes, etc.?²

What I mean by 'observable' here is just what is accessible to the unaided human senses. The word 'observable' is like 'breakable' and 'portable'. I would not call this building or a train locomotive breakable just because we now have instruments that can break them – nor call a battle tank portable because it can be carried using a Hercules transport plane. In the same way the word 'observable' does not extend to what is purportedly detected by means of instruments.

But this opens up the immediate second challenge: we humans change too, not just our technology.³ Evolution has not stopped, who knows what we can yet grow into? A good point, and that is what I want to take up today.

II. Observability perspectival

The first point I want to make is that 'observable' is redundantly equivalent to 'observable by us' – in that way too it is like 'breakable' and 'portable'. And yes, we can change. So thereby hangs a tale ...

Observation and the epistemic community

The 'able' in 'observable' does not look indexical. But it is; it refers to us – just as it does in 'portable', 'breakable', and 'potable', though perhaps not in 'computable', and certainly not in 'trissyllable'. Just now it does not seem to make too much difference whether we say 'within our limitations' or 'within human limitations'. But the range and nature of beings that we count as *us* is not fixed, either necessarily or even historically.

The observable phenomena consist exactly of those things, events, and processes that are observable to us. We may very well be quite certain of who we are, and may have full beliefs about the characteristics that all and only we have in common. Suppose those common features are summed up in 'human.' Then we fully believe that the observable phenomena are exactly the humanly observable phenomena. But we realize that we are in evolution. We are also not so given to tribalism or species-chauvinism as to see those common features as essential. Even a modal realist could say 'We are human, and humans are essentially X, so each of us is essentially X, but we may (could, might) in the future have beings among us who are not X.'

Epistemology and cognitive science part ways here. The cognitive scientist, in so far as s/he engages in empirical research and not merely in theorizing, studies human and animal information processing. In epistemology we must take the indexical seriously and reflect also on how we are to think of our own beliefs, opinions, and epistemic activity in general in the light of those contemplated futures where *we* and *human* do not coincide.

The Day of the Dolphins

The touchstone for the difference will clearly appear when we envision changes in the epistemic community, with consequent changes in the referent of 'we', 'us' and 'our.' This was noticed at once by various scientific realists who tried to exploit it in arguments against constructive empiricism. It will be instructive to diagnose the fallacies in those arguments for that will in fact lead us to genuine problems for anti-realist epistemology.

Smart detectors and bionic persons

Humans equipped with instruments, surgically implanted electronic devices, evolutionary stories of progeny who grow electron microscope eyes or eventual assimilation of dolphins or extraterrestrials into our community ... these all are ways in which we, and our self-conception, could change. In such changes, what is *observable by us* also changes. *Question:* does that not change right now what we can give as reasonable and intelligible content to 'observable'?

Let us carefully consider the form of argument that challenges, by such illustrations, the observable/unobservable boundary on which constructive empiricism relies:

- 1) We could be or could become X.
- 2) If we were X then we could observe Y.
- 3) In fact, we are under certain realizable conditions, like X in all relevant respects.
- 4) What we could under realizable conditions observe is observable.

Therefore:

Y is observable.

Certainly a valid argument. But what does it look like when instantiated to a particular content?

Suppose we take as our example the alpha particles familiar from the earliest descriptions of cloud chambers. Let Y be alpha particles. Let X be an

The Day of the Dolphins

organism with special senses that operate like the cloud chamber, plus sensor that registers what corresponds to cloud chamber tracks. Then premise 2 says:
If we were (or were like) X, then we could observe alpha particles

What is the basis for this claim? It is clearly a theory which implies that those tracks in the cloud chamber are made by alpha particles. And they are indeed made by alpha particles if alpha particles are real and that theory is true. But our current acceptance of that theory – even if we in fact accept it – does not imply belief in that much. So the argument already assumes, for its polemical success, a different construal of acceptance of scientific theories, thus begging the question against constructive empiricism.

Just to round out the picture: the same sort of question begging presumption surrounds premise 3, when given concrete content. We do not really need to appeal to AI, current electronics, or new physics, let alone molecular biology to provide the setting for the arguments. We only need the indubitable possibility of a smart detector of, for example, single electrons or single photons or single alpha particles. But that possibility is already easily within reach, if anyone cared to adapt the technology – always, however, on the assumption of the reality of those particles and truth of the relevant theory. For rather than waiting for the emergence of new kinds of smart detector, we can modify one we know already – me or you, for example.

To do this we rig up a physical detector, coupled to an amplifier, which can register the impact or presence of a single electron. It emits the sound 'Bingo' whenever that happens. If we point an electron gun at it, designed via our current theory to emit one electron per minute, the apparatus emits the sound 'Bingo' at that rate too, and so forth.⁴ Now we detach the loudspeaker, and link the output, perhaps a little less amplified, to an electrode in someone's brain. The output change is reliably detected by him in the form of an indefinable feeling, perhaps only scarcely at the level of consciousness. Nevertheless this is sufficient

The Day of the Dolphins

for the next step – we condition him, by means of biofeedback techniques, to shout ‘Bingo’ when he gets that feeling.

Now we have our smart detector of single electrons, and he was accepted as a member of *us* already. He can carry the apparatus strapped to his back; we can also bring about the existence of whole regiments of such smart detectors. And we can reliably evoke the shout ‘Bingo’ by coupling on an electron gun and pulling the trigger. So should we say right now that single electrons are observable?

In fact, now that we have realized the crucial experiment so easily, it is not so convincing anymore. If we bracket the theory involved in such terms as ‘electron gun’, we have simply a sequence of now (already) observable events with reliable predictions. And we realize actually that not only did we not need faith in AI, but we do not even need electrodes on the brain. For the relevant possibility was already there in Schroedinger's famous remark that the emission of a single photon can sink a battle ship. All it needs is an amplifier whose output is coupled to an Exocet missile launcher. It would not take biofeedback techniques to train someone to shout ‘Bingo’ every time the apparatus sinks a battleship. And here again we have a reliable smart detector of single photon emissions in a specially arranged suitable context (that is, the experimental arrangement requires a randomly operated off-on switch on the apparatus and suitably positioned battleship; he is trained to shout exactly if he sees a battleship sunk under those conditions. He will be very reliable even if there are a few other missiles flying around in the area.)

But now of course the possible existence of potential believers who, according to our theory, are reliable single electron or photon detectors, no longer looks like it could establish very much. The reason is that we have as usual produced a situation in which our predictions, even by us non-rigged-up people, are reliable and concern observable events in the present sense. That is we can

The Day of the Dolphins

reliably predict that pulling the trigger on a macroscopic object we have classified as an electron gun in good working order will be shortly followed under the described circumstances by a shout of 'Bingo' from the first experimental subject. And we can reliably retrodict the position of the randomly operated switch, from the second subject's 'Bingo' shouts.

Changes in our epistemic community

The challenge in terms of humans who grow electron microscope eyes and the like takes also a second form, due to William Seager.⁵ This relates specifically to how we should think about our own epistemic future when we contemplate the widening of our epistemic community – the 'us' in '*observable by us*'. In doing so we are contemplating the end of the equation of *observable by us* with *humanly observable*, at least in the sense in which 'humanly' refers to the sorts of animals that we early 21st century humans are. This widening could bring into our community dolphins, extra-terrestrials, or the children of *Childhood's End*. Here is the challenge. Let us suppose that I now admit some positive probability for the admission – at some future date – of dolphins as persons, as bona fide members of our epistemic community. Suppose furthermore that I currently *accept* (but do not *believe to be true*, only *believe to be empirically adequate*) a science which entails that dolphins are reliable detectors of the presence of Ys. Here Ys are things that I currently classify as unobservable, since they are not detectable by us even if they are real. Add for good measure that at present we are 'atheistic' in this respect and believe that Ys are not real! Now it could be part of the supposition that dolphins themselves will claim evidence which refutes that present science. Let us not suppose that! Let us make it part of the story that after this widening of the community we shall still accept the theory. So at that point we will add: 'Some of us observe Ys.' We will add by implication that Ys are real.

The Day of the Dolphins

There is no great threat in the reflection that in the future we shall give up some beliefs we hold now and replace them by contrary beliefs. But this is a special case, and we can spell out the argument as follows:

- 1) the science I accept as empirically adequate entails that Ys exist
- 2) the science I accept also entails that in some possible future our community will include members who observe Ys

Therefore:

- 3) I should now believe that Ys exist.

Seager offers the following as an analogy to our worrying situation with respect to dolphins in the above case:

- 1*) We know that if we encountered rational creatures of type X who sincerely informed us that the earth would explode tomorrow, we would believe that the earth would explode tomorrow.
- 2*) We know that rational creatures of type X are possible.
- 3*) We know that if we were to encounter rational creatures of type X, they would in fact sincerely inform us that the earth will explode tomorrow.

Therefore:

We should now believe that the earth will explode tomorrow.

The crucial supposition behind 1*) is that we accept a theory which entails that certain observable events (communications from the Xs) are reliable indicators of earth explosions (also observable events) to come. Thus we are appealing to the empirical adequacy of our background theory only.

We have here in fact a good analogy for the dolphins, except for the current observability of the explosion: the Ys were not currently classified as observable. That introduces a disanalogy also for 2) and 2*): in the case of 2*), our acceptance of the background theory as empirically adequate leads us to a positive probability for the existence of reliable observers who can predict earth explosions.

The Day of the Dolphins

To display the equivocation in the original argument we need only emphasize the indexical character of 'observable'. In the envisaged scenario we see ourselves as truthfully saying at a certain later time, in a certain possible future:

(A) Some of us are observing Ys; thus, Ys are observable

But it would be a mistake to infer anything like:

(B) Ys will be observable at that time

For in (A) we have a sentence uttered truthfully at a later time, while (B) would be our statement now. The fallacy involved is the same as in:

When we are in Pisa next July we will truthfully say 'The Leaning Tower is here'.

Therefore:

The Leaning Tower will be here in July

As I pointed out to begin, 'observable' does not look indexical, but it is. That there are these hidden, subliminal indexical aspects to some of our discourse is precisely the lesson we learned from Putnam's Paradox. However, Seager has in effect pointed us to a deeper problem that will provide us with a greater challenge.

The new riddle of prevision

The problem Seager posed is not so easily dismissed, for it carries a strong intuitive sense of puzzlement. What if we do experience such a change in what counts as us, as our epistemic community, and thereby see a profound sea-change in our relation to nature as a whole? Just how are we to conceive of ourselves as epistemic continuants, once we contemplate such radical changes in the range of epistemically accessible aspects of nature in our future?

The question will become crucial at a particular point in the story: the point where we realize that the ceremony of admitting the dolphins to our community is about to be performed. Suppose it has not been performed yet, but we already know it will be soon – at that point we must surely change our mind

The Day of the Dolphins

about the Ys already? Let's go back a bit further: the ceremony has not been decided upon, admission to membership is still being debated, but has become very likely. At that point should we not at least let go of our out and out agnosticism or 'atheism' and admit that it is very likely that our science is also right about some unobservable parts of nature? Now, going back still further to our present position, when all we have is the theoretical possibility, should we not think that such future likelihood, if it is to arrive, will have grown from a small likelihood in our minds already? Or rather, a small likelihood that should already be there for us? In which case already now, before we have seriously encountered those dolphins or whatever as yet, we should not be completely agnostic about our science's truth.

I could continue to block this rhetoric by insisting that there is a gulf of principle between possibility and positive probability. But the gulf cannot be one uncrossable by a belief or rational positive probability that we will in fact admit such creatures. We need to look more deeply into general epistemology. How shall we contemplate the possibilities of our own future opinions subject to such changes? As guiding analogy of a much more general sort I want to ask about what I shall call *epistemic marriage*.

Epistemic marriage

Consider the following conception of marriage. After the wedding, the two constitute a couple, and there is no longer personal but only communal opinion for them on all subjects to which both have equal access, including access through the partner's reports of private experience (admitted on equal footing with memory of one's own private experience). This is similar to the dolphins problem, except that the envisaged union is more intimate (and is clearly a matter of decision, not easily seen as forced by opinions about what things, people, etc. are in fact like).

The Day of the Dolphins

The question this raises is whether I can beforehand envisage such a change, while maintaining that I shall always change my opinions and beliefs *rationally*. Perhaps the constraints of rationality on opinion do not affect much else. For example, they may give me no problem at all if I foresee perfect epistemic domination by myself of my partner – that s/he will submit him or herself epistemically to my dominion. But then the question is whether that transition will count as rational opinion-management for my partner, and one doesn't suppose that general epistemology is a respecter of persons.

Consider then my present status before entering into such a union; suppose I consider it part of one of my possible futures. There are various possibilities for the type of person I shall marry, but some of them may have drastically different opinions from my own. Should I therefore expect that the couple of which I shall become part through fusion of this sort, will have opinions drastically different from my own? Does that not mean that my present opinions must be accordingly 'diluted' – in the worst case, in which I foresee having to give up probability *one* or *zero* especially, it has to be given up right now?

Example: I give probability *zero* to reincarnation. Do I now face the dilemma of either giving it a small positive probability, or else rejecting as certainly false the supposition that I shall marry a believer in reincarnation?

To use another example, suppose I am considering as spouse someone who claims to be psychic, and I presumably do not believe in that.

I don't think I shall now say that I shall not marry someone whose beliefs disagree in some way with mine. I can also foresee that after the marriage, or at that time, or as part of the decision to marry, I shall acquire these contrary beliefs. But at the same time I need not hold that I endorse this as the way to acquire reliable opinions. *In other words, I foresee a break of epistemic integrity.* This is so even if (a) I am presently agnostic about psychic powers and (b) I believe that no

phenomena which are observable to *me* will disprove the reality of psychic powers.

This is a disturbing reflection. My epistemic integrity is compromised when I allow that I shall possibly go along with something like this. Realizing that the married couple will legally inherit all my present contracts and obligations, it appears that I am ready to incur a certain loss for the me-couple. It is a bit like what Americans call marriage tax, only worse.⁶ Clearly we have reached a fundamental difficulty in our epistemology, and we need to go back to fundamentals. We have to examine the terms of our discussion – opinion, belief, and the constraints of rationality on how we manage them – on a more fundamental level.

III. Self-Prevision: its logical laws, its subjective sources

So the topic we have to investigate is what might be called our reflective opinion: this includes both how we view our current opinions and how we envisage what they may be in the future.

Hintikka's problem

As a start I want to remind you of a mistake in those heady days when modal logic seemed to provide a royal road to philosophical enlightenment. There was a logic of everything so of course there was a logic of belief. The seminal text was Jaakko Hintikka's *Knowledge and Belief*.⁷ What is the logic of belief? That means: what inferences about belief are valid? Consider:

X thinks that A; *therefore* X thinks that B

This relationship of entailment can presumably be captured in a logical system, the logic of belief. Unfortunately that entailment relation was trivialized by what we might call 'the problem of the moron'. Whatever sentences A and B are, no matter how closely logically related, there was a conceivable person of

sufficiently low logical acumen who wouldn't get it. So we need a new approach to the logic of belief and opinion.⁸

Logic in the first person

On purely logical grounds we can see that if someone is of the opinion that A that may bring with it a commitment to or responsibility for B, on pain of incoherence. The paradigm example here is Moore's Paradox. If I were to say

It is snowing and I do not think that it is snowing

then I would display an incoherence in my state of opinion. I cannot say this, but not because it could not be true – I cannot say it on pain of incoherence. We can describe this fact about my opinion in terms of an inferential relationship:

It is snowing >> I think that it is snowing

I think that it is snowing >> It is snowing

But both in its meaning and in the logical laws obeyed this is quite different from the standard logic, for we certainly would not infer from the above that

IF I think that it is snowing THEN it is snowing

is a valid sentence.⁹

The first person character of these sentences is of course crucial to this relationship. There is nothing wrong with 'It is snowing and Paul does not think that it is snowing'. Indeed, there is nothing wrong with 'It WAS snowing and I DID not think that it WAS snowing' or 'There will be times when it WILL BE snowing and I SHALL NOT think that it is snowing.' The reason would appear to be that 'think' in the first person present tense has the linguistic function of *expressing* my opinion. The difference between stating what our emotions, values, and intentions are on the one hand and expressing them on the other is of course familiar. That contrast is crucial also for opinion. Secondly, we can express our opinion only in indexical, self-attributing fashion. Opinion is perspectival.

There is a possible ambiguity here. In a therapy session a person may perhaps come to the realization of a surprising autobiographical fact: he discovers what his opinion really is. In that context, even ‘I think’ may play the other, simple fact-stating role. So I propose that for our present inquiry it is best to make a syntactic distinction that we do not see in English, and I will italicize the words when they play the expressive role, and use bold face for ‘think’ in the fact-stating role.

It is snowing >> *I think* that it is snowing

I think that it is snowing >> It is snowing

I think that it is snowing >> *I think* that **I think** it is snowing

are all valid. But

I think that it is snowing >> *I think* that *I think* it is snowing

makes no sense; the expressive use does not iterate (the fact-stating use can).

However we should note also that the ‘*I think*’ accompanies every thought, as Kant said; and for this very reason we can usually let it go without saying. So in fact in ordinary discourse it is often left out.

Subjective probability

We need now to complicate the picture just a bit by allowing for more nuances of opinion. Mostly I don't just believe that this or that will happen – it only seems more or less likely to me. In our examples, we already allowed for this. Now let's make it official by replacing that ubiquitous *I think* with *It seems ... likely to me that*, with the blank filled in with some degree. Do not think numbers right away: our opinion is typically too vague for that, though we have natural ways of being more precise. Compare:

It seems likely to me that it will snow

It seems very [extremely] likely to me that it will snow

It seems twice as likely to me that it will snow than that it will rain

The Day of the Dolphins

It seems twice as likely to me as not that it will snow

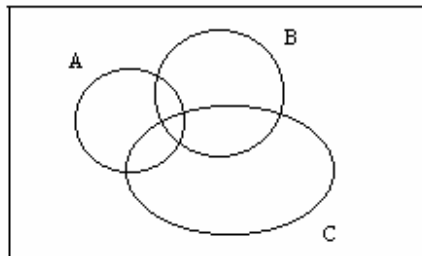
The last one is numerically precise, it translates directly into:

My subjective probability that it will snow is 2/3

Symbolically: $P(\text{it will snow}) = 2/3$

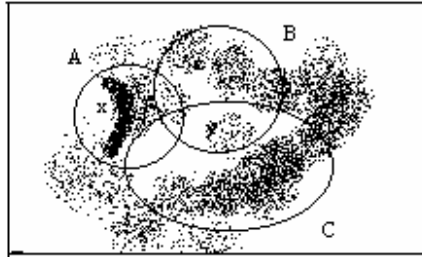
To accommodate the vaguer examples, we must allow for intervals like $[0.1, 0.4]$ to replace the number.

Before going on I want to give us an intuitive grasp on how we do actually reason in this format. My full beliefs together give me one picture of the world, not a very complete one obviously – but that is where I say ‘*That is what things are like!*’ About all the alternatives left open by these full beliefs, though, I am not so definite: that is where I say those sorts of qualified things illustrated above. Now, one way to keep this scheme before our eyes is by means of what I call the Muddy Venn Diagram. Just as in elementary logic class we depict the space of all possibilities by means of a Venn Diagram:



But then we smear and heap mud on it, to indicate proportionally how much credence we give to these various possibilities:

The Day of the Dolphins



Suppose that the total is 1 kg. of mud; then if A has 1/3 kg on it, that indicates that my probability for A being the case is 1/3. This is not just a mnemonic: using this we can immediately see what the logical rules for consistency must be. For example, just thinking about the amount of mud on the various parts, we can see that:

$$P(A \text{ and } B) + P(A \text{ or } B) = P(A) + P(B)$$

which is the Axiom of Additivity for probabilities.

Notice also that to teach ourselves to reason with vague probabilities we should just learn how to do it with precise ones. When children learn how to deal with such judgments as *John is about 5'9' and Julia about 6'2'* they do not need to study a special calculus of approximate numbers – their school arithmetic is all they need to understand that Julia is taller than John. Similarly with our vague degrees of belief.

Opinions can be stated as well as expressed of course. We can also make state attributions saying that so and so has some such epistemic attitude, to describe his or her opinion (in part). To revamp some of our previous examples: *It seems likely to me that it seems unlikely to Jeremy that* it will snow.

We must allow for both precise and vague probability here. Here is an example with several of the above features:

$$P(p_{\text{Jeremy}}(\text{it will snow})) = [.5, .75] = .8$$

The Day of the Dolphins

The initial 'P' needs no subscript for it expresses always the speaker's current opinion. The bold face is again to be used for biographical and autobiographical statements of fact.

To continue our discussion of reflective opinion, we should now ask about expressions of form:

$$P(\mathbf{p}_{ME}(\text{it will snow})= \dots) = \text{---}$$

Until I have spent some time with that therapist I may not be too sure of what I think, so this makes sense. But in view of the sorts of logical relations we saw above, what are the constraints of coherence for some thing like that?

Let's ask concretely to what extent I could coherently express some lack of confidence in my own opinion. What of the possibility that some proposition A that seems unlikely to me is in fact true? How likely does that seem to me? You understand that we are in Moore Paradox land here. Coherence requires precisely that if we say something of this form:

$$P(\text{It will snow and } \mathbf{p}_{ME}(\text{it will snow})= x) = y$$

then the number y must be no greater than x. So here too we have a significant logical point although solely about what expressions of opinion will and will not display an incoherence in that opinion.

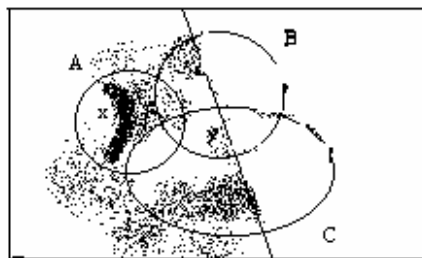
Rational change in view

I'll stick with the fiction of sharp, numerical probabilities for now, and leave the more realistic (hence more complicated) presentation for another time. The simplest case of a change in opinion is the one where some newly acquired bit of belief triggers *modus ponens*. For example, I come into the kitchen and I see small black droppings and note bite marks on the cheese. If I immediately conclude that we have a mouse, some people think that I have made an inference to the best explanation. But I had no need of any such move or maneuver: I

The Day of the Dolphins

already thought all along that if there are such changes in the kitchen scene then it was visited by a mouse.

That is just *Modus Ponens*, you'll note. If I am not quite as dogmatic in my beliefs then this evidence will not take me that far, but I will just go to a pretty high probability that there are mice. We can again make this visually intuitive with the Muddy Venn diagram. The space of possibilities before I came down the stairs was divided into two parts: the part that has my kitchen with small black droppings and bite marks on the cheese, and the part that does not. When I see



the evidence I simply wipe off all mud from the ruled out part:
This is called

CONDITIONALIZATION, the probability analogue of *Modus Ponens*.

But even this simple logical updating is already a bit more complicated with degrees of belief. First of all, I may only wish to raise my degree of belief that there was a mouse, not raise it to certainty. Secondly, if I raise that, I cannot leave all the rest alone, for that will affect my views on how our house relates to the local fauna in general. Enter here the probability calculus: it is the logic that spells out what coherence requires on my opinion in general. It even gives us at least the resources for describing what purely logical updating in response to new evidence can be like.

Enter here also a major epistemological rift. The old fashioned idea that we must proportion our belief directly to the evidence – as propounded by Locke and oft

The Day of the Dolphins

repeated since – has as its descendant Orthodox Bayesian epistemology. This position implies:

Purely logical updating in response to new evidence – i.e. **Conditionalization** – is the sole rational form of changes of opinion.

You can see at once that this position will rule out epistemic marriage in all but the trivial cases.

There are more liberal alternatives. The Bayesian will say that if you are willing to depart from pure liberal updating, then *anything goes*. That is not so; we can again insist on coherence constraints, *at least in our present opinion about the future opinions we may come to*. Specifically, I consider the following to be mandatory for present coherence¹⁰:

General Reflection Principle: My current opinion about event E must lie in the range spanned by the possible opinions I may come to have about E at later time t, as far as my present opinion is concerned.¹¹

As an example, think of what would violate this principle. You are going to buy a lottery ticket, and I ask you: if the number ends in a 0, will you think that you are likely to win something – and you say NO. The same for my questions about 9, 8, ..., 1. After all that you still say ‘But I am feeling lucky! I will buy the ticket because I think I am likely to win this time!’ Well, that violates the Principle.

One important consequence of this principle occurs immediately if we apply it to numerically precise subjective probability. That is the ‘ordinary’ or ‘simple’ Reflection Principle:

$P(\text{It will snow, given that } p_{ME[t]}(\text{it will snow})=x) = x$

For example: On the supposition that an hour from now it will actually seem K times as likely to me as not that it will snow, it does seem K times as likely as not to me that it will snow later today.

The Day of the Dolphins

where the autobiographical attribution now concerns my opinion at some later time t , and t here is a relative time (such as 'tomorrow' or 'later today' or 'one hour from now'). How can this be deduced? We have to read the General Principle as applying not only to probability but to *expectation* in general. By this I mean that opinion has as most general form in this context that of:

My expectation of my salary increase is 4%, because it is equally likely that I will be evaluated as deserving a bonus (in which case the increase will be 6%) or as deserving no bonus (in which case it is 2%)

My salary increase is a quantity which can take various possible values, and my expectation of it is my subjective average for the possible scenarios I can envisage. Then the trick is to treat one's own future opinion as such a quantity. Applying the General Reflection Principle to that quantity will then yield the 'simple' Reflection Principle.

But this 'simple' Principle has looked very suspicious to many people (even though the orthodox Bayesian clearly satisfies it, and has not for that reason looked suspicious to anyone!) so we should make sure it does not say too much.

This Principle does not make it impossible to express either confidence or lack of confidence in my future opinions, but not in one direction or other. I may expect that my future probability for something will be out by some factor, either too low or too high, but not be sure that it will be too low, nor be sure that it will be too high. The Principle does entail a more general form also of the nuanced Moore Paradox point. I can certainly say¹²:

$$P(\text{It will snow and } p_{ME[t]}(\text{it will snow})=x) = y$$

For example: It seems N times as likely as not to me that the following are both true: It will snow later today and it actually seems K times as likely to me as not that it will snow.

The Day of the Dolphins

But y must be no greater than x ! If y is greater than x (or N than K) then the Reflection Principle is violated.

One uncompromising limit, however: if I am now sure that I will have a certain opinion in the future, then I must have it now – on pain of present incoherence.

Wesley Salmon mentioned someone's comment on the principles of Scientology: 'I am an empirical scientist so I won't say they are false before the evidence is in. But when it is I will!'

What does this anecdote illustrate? If this scientist's opinion is coherent then of course he has signaled that he already disbelieves those principles. And he does so because he already fully believes that the evidence will go one way and not another.

I realize that this Principle makes no sense if we simply see it as concerning factual prediction. There may come to be serious physiological and psychological deficiencies in my future. But if I express an opinion that violates the General Reflection Principle then I display a deficiency either in my current opinion or else in the way I shall go about managing my opinion in the future. As analogy imagine me saying: 'There will be arithmetical mistakes in my budgeting for next year'. The problem is not that this sentence cannot be true – it can – but that I am expressing something that violates norms I should be expected to uphold. We want to reply 'So do something about it!' – and that is just what the General Reflection Principle signifies.

Finally, how much weaker is this Principle than the orthodox Bayesians' insistence that the right rule is always to conditionalize? The two coincide precisely when the person is sure that s/he can canvass all the possible outcomes, and say what his/her posterior probability would be in each of those cases.¹³ So the Reflection Principle has all the bite there is to be had for exactly those people who cannot foresee how they will make up their minds under all possible circumstances. Not exactly an implausibly conjured class!

IV. Epistemic marriage revisited¹⁴

Recall our concept of epistemic marriage. After the wedding, the two constitute a couple, and there is no longer personal but only communal opinion for them on all subjects to which both have equal access, including access through the partner's reports of private experience (admitted on equal footing with memory of one's own private experience). The 'marriage' could be assimilation of dolphins or extraterrestrials into our epistemic community as full and equal members.

In marriage one hopes for a certain degree of symmetry and equality as well as harmony. This is also what various studies have looked for in the pooling of opinions and preferences. We can begin modestly by suggesting that any views already held in common should be preserved in forming the views of the unit. Such conditions are called *Pareto conditions*, and can take various forms. Let the partners be X and Y, forming the unit U:

[P₁] If A and B seem equally likely to both X and Y, then A and B are to seem equally likely to U.

We arrive at conditions [P₂] and [P₃] by replacing 'A and B seem equally likely' by 'A seems at least as likely as B' and 'A seems more likely than B' respectively, adjusting *mutatis mutandis*. A stronger condition is this:

[P₄] If A seems at least as likely as B to either X or Y, and seems more likely to the other, than A is to seem more likely than B to U.

These conditions can all be satisfied, provided X and Y have coherent states of opinion to begin. In fact it is quite easy to see how: They simply settle for a degree of belief somewhere between the initial two. To do this systematically so that the result will be coherent they choose a linear combination:

If they have sharp probability functions p and p' then U can be given any *mixture* of these, that is combination $q = ap + (1-a)p'$.

If in addition, as in any good marriage, we require symmetry, the proper combination would be half and half: $(1/2)p + (1/2)p'$.

The Day of the Dolphins

So it can be done.¹⁵ But will it count as rational? We have here precisely one of those episodes in which we feel ourselves called to account for the change in view.

Imagine that I marry someone who believes in reincarnation, and we settle on the common opinion that reincarnation is precisely as likely to be real as not real. If she was initially [almost] certain about it, and I [almost] certain of the opposite, this would be half-way between. But did I conditionalize on new evidence? Something new has happened—the marriage. But did I have the prior opinion Reincarnation is as likely as not to be real, on the supposition that I marry someone who believes in it which would be required for conditionalizing to lead me to the new opinion?

Most likely not.¹⁶ So after all this we seem to have only succeeded in making our problem clear. While there seems to be a way to form the epistemic union, there seems to be no way to rationalize the consequent individual change of opinion, along traditional lines.

But it is not irrational to be so struck by the appearance of rival opinions to one's own. Indeed it seems rational to accept the appearance of such a rival as requiring an attempt to stand back somewhat from one's own point of view. The proper response would seem to be something like this: stand back, bracket the differences between the two, and then let the resulting 'neutral' opinion evolve to a less neutral one in response to the evidence. Mixing may be an attempt to do something of that sort, but it miscarries for it actually results in a sharply discontinuous change of opinion – prejudging what the outcome should be. Nevertheless, there must logically be many different ways to do this. Are there any constraints on this?

The Reflection Principle applied

Even just given the General Reflection Principle, *mixing* is also not an acceptable prospect. For think of any situation once I have decided on this marriage, and the ceremony is about to begin. At this point I am [almost] certain that reincarnation is not real, yet certain that very soon now I shall instead judge it to be as likely as not. That violates Reflection.

But we can imagine slightly different situations in which Reflection is not violated. Imagine that there are two partners I may marry: one strongly believes in reincarnation and the other strongly disbelieves. I foresee that if I marry the one, my degree of belief in reincarnation will go down, and if I marry the other it will go up. Since I do not privilege one direction over another I may satisfy Reflection. This works only as long as I remain suspended. There is also another way in which I may satisfy it even if I can marry only one of these two. Suppose the epistemic unit is formed the moment it is decided upon – and suppose the decision will be unforeseen and unsettled until the very last moment. Then again there may be no incoherence. This solution we could call that of 'epistemic elopement'. *Reflection may be unviolated in the case of sudden unpredictable epistemic elopement.*

To envaguen

However the fact remains that foreseeing that one's opinion will change to a mixture of one's current opinion with that of another is a clear violation of Reflection. Prospects for epistemic marriage seem dim. But in fact there is a solution. It won't help those who insist on conditionalization, but will satisfy the Pareto conditions and Reflection. The solution for the partners is not to settle on a specific spot in between, but to *envaguen* (to make their opinion vaguer).¹⁷

The Day of the Dolphins

This is easiest to illustrate if they begin with sharp subjective probabilities. Suppose that I and my partner have probabilities 0.01 and 0.99 for reincarnation. Rather than settle on 0.5 we agree that

The probability of reincarnation is no less than 0.01 and no more than 0.99

shall encapsulate our *entire* assessment of how likely reincarnation is. (If X and Y have probability functions p and p' then U will have the function P : $P(A) = [p(A), p'(A)]$.)

It can, and should, be part of their commitment that they will have a common *epistemic policy* as well. Both can then hope that as they follow that policy to manage, amend, and update their common opinion, it will converge on the prior opinion s/he brought to the marriage.

How will this affect the dolphins problem? Before union we do not think that dolphins observe Ys, when we are still agnostic about whether Ys are real at all. After the union, our common opinion will be at least as vague on the matter as either of us was. Together we will go over the evidence, once we are truly both contained in the 'us' of 'observable to us.' What will happen? We can't say without violating Reflection.

Conclusion

Let me quickly recapitulate and draw a moral. Real anti-realism must be a position that can only be expressed in the first person. (Preferably the first person plural ...) But that will be no more than an empty sound if we don't then also exploit that to change our understanding of traditional philosophical problems.

It would be a great boon for epistemology if it got itself definitively out of the 'X knows that p' mess as well as the skepticism syndrome and all other such sado-masochistic dead horse entanglements. But there are two ways out, one illusory and the other fruitful. The illusory one is what Quine called naturalized epistemology. Certainly, philosophers should study scientific models of

The Day of the Dolphins

information-processing, especially in physics and computer science. But these models represent only the physical correlate of the epistemic process. If we simply transpose them to the human case we are forgetting that the basic philosophical questions apply to the sciences as well. To be a good way out of the past the way out needs to do justice to the past and to recapture what was valuable in it even while rejecting it.

So, in going back to the topic of observability I wanted not only to solve an outstanding puzzle but also to illustrate the good way out. That way is to ask what a philosophical problem looks like once we really put ourselves back into the picture.

You may not immediately have appreciated this when I brought in subjective probability. I won't blame you if you thought 'Been there, seen that, had enough of it!' For this subject was another one treated with great technical virtuosity together with such a lack of critical concern with traditional philosophical issues that I cannot blame you. There has been a sort of subjective probability slum in philosophy, and its inhabitants, me included, have not convinced many other philosophers that what happens there is anything more than technical self-indulgence.

But I think this will change if subjective probability is put in the first person, and its problems recast at a fundamental philosophical level. For then it will become clear that we have there, however imperfectly still, a way of representing opinion that shows up the naiveté and oversimplification inherent in much of traditional epistemology.

I've meant to comment on the day of the dolphins as only one example of how a philosophical question may be transformed when we switch in descriptive epistemology from the simple trichotomy of belief/disbelief/neutrality to subjective probability as our framework. I submit that there will be a similar transformation of other philosophical questions if approached in this way,

The Day of the Dolphins

creating in each case a new array of problems and puzzles to be addressed, solved, dissolved, or shown up as further illusions of reason.

NOTES

- 1 I happily dedicate this essay to my friend John Woods who has, ever since our Toronto days together, inspired me with his sustained inquiry into the mysteries of reason, both formal and informal.
- 2 See Hacking, 'Do we see through a microscope?' in *Images of Science: Essays on Realism and Empiricism, with a Reply by Bas C. van Fraassen*, eds. Paul M. Churchland and Clifford A. Hooker (Chicago: University of Chicago Press, 1985), 132–52, and my reply in the same volume.
- 3 See especially Paul Churchland, 'The ontological status of observables: in praise of the superempirical virtues,' 35–47 in Churchland and Hooker.
- 4 A TV or computer monitor basically consists of an electron gun at one end with a sensitive screen on the other. The computer can read the screen, so respond if a particular spot lights up, and emit a sound. Not exactly advanced technology today, except for the degree of sensitivity we are imagining here.
- 5 William Seager, 'Scientific anti-realism and the epistemic community,' in *PSA 1988, Vol. 1: Proceedings of the 1988 Meeting of the Philosophy of Science Association* eds. A. Fine and J. Leplin (Philosophy of Science Association, 1988), 181–87. This was part of a symposium on Realism at the Philosophy of Science Association Biannual Conference of 1988, in which I acted as commentator.
- 6 When two people with incomes marry, their joint income can go into a higher bracket, with higher taxation rate.
- 6 Jaakko Hintikka, *Knowledge and Belief: an introduction to the logic of the two notions* (Ithaca, NY: Cornell University Press, 1962).
- 8 The approach here outlined is that of Chapter 7 of my *Laws and Symmetry* (Oxford: Oxford University Press, 1989).
- 9 It is to be remarked that the same applies to *It is true that* if the language has truth value.
- 10 See my 'Belief and the problem of Ulysses and the Sirens,' *Philosophical Studies* 77 (1995): 7–37. The less general Reflection Principle also noted below I had introduced in 'Belief and the Will,' *Journal of Philosophy* 81 (1984): 235–56.
- 11 'Opinion' here covers both probability and expectation. Semantic and set-theoretic paradoxes threaten if such a principle is left with the range of applicability unrestricted.
- 12 Note that unless $x = 1$, I cannot conditionalize on the statement [It will snow and \mathbf{p} (it will snow)= x], with \mathbf{p} indexed to myself now; for if I gave it probability 1 then I would be in violation of the Reflection Principle. But the content of that statement, equally expressed by some eternal sentence of course, is a proposition which I could give probability 1 at some other time.
- 13 See my 'Conditionalization, a New Argument for,' *Topoi* 18 (1999): 93–6.

14 In this section I shall be implicitly referring especially to two papers: Teddy Seidenfeld and Joseph Kadane, 'On the shared preferences of two Bayesian decision makers', *Journal of Philosophy* 86 (1989): 225–44, and Philippe Mongin, 'Consistent Bayesian aggregation,' *Journal of Economic Theory* 66 (1995): 313–51.

15 More stringent Pareto conditions are not as easy to satisfy; and if preferences are to be balanced as well as probabilities, we run into unsolvable problems. See the papers cited in the preceding note. Moreover, even if we leave values and preferences out of account, there is a problem about preserving agreed on correlations, due to Simpson's paradox; see e.g. *Laws and Symmetry*, 204–5.

16 If we made such prior opinions a requirement for rational epistemic union, the dolphin problem would not be problematic either. For then we would not accept them unless already *beforehand* we had concluded that whatever they called observable was in fact observable. That would mean: 'What they say is observable to them is observable to us' pronounced at the earlier time when 'us' still excludes them.

17 Vague probability is itself a topic with much technical literature and remaining problems. See for example Richard Jeffrey, 'Bayesianism with a human face,' in *Testing Scientific Theories*, ed. J. Earman (Minneapolis: University of Minnesota, 1984), 133–56; my 'Figures in a probability landscape,' in *Truth or Consequences*, eds. M. Dunn and A. Gupta (Dordrecht: Kluwer, 1990), 345–56; and Joseph Y. Halpern and Riccardo Pucella, 'A logic for reasoning about upper probabilities,' To appear, *Proceedings of the Seventeenth Conference on Uncertainty in AI*, 2001. (Accessible at: <http://www.cs.cornell.edu/home/halpern/papers/up.pdf>)