3 Probabilism

3.1 Justifying the probability axioms

The hypothesis that rational degrees of belief satisfy the mathematical conditions on a probability measure is known as **probabilism**. In this chapter, we will look at some arguments for probabilism. We do so not because the hypothesis is especially controversial (by philosophy standards, it is not), but because it is instructive to reflect on how one could argue for an assumption like this, and also because the task will bring us back to a more fundamental question: what it means to say that an agent has such-and-such degrees of belief in the first place.

We will assume without argument that rational degrees of belief satisfy the Booleanism condition from p.24. The remaining question is whether they should satisfy Kolmogorov's axioms (i)–(iii):

(i) For any proposition A, $0 \leq \mathrm{Cr}(A) \leq 1$.

(ii) If A is logically necessary, then $\mathrm{Cr}(A) = 1$.

(iii) If A and B are logically incompatible, then $\mathrm{Cr}(A \vee B) = \mathrm{Cr}(A) + \mathrm{Cr}(B)$.

Consider axiom (i). Why should rational degrees of belief always fall in the range between 0 and 1? Why would it be irrational to believe some proposition to degree 7? The question is hard to answer unless we have some idea of what it would mean to believe a proposition to degree 7.

A natural thought is that axiom (i) does not express a substantive norm of rationality, but a convention of representation. We have decided to represent strength of belief by numbers between 0 and 1, where 1 means absolute certainty. We could just as well have decided to use numbers between 0 and 100, or between -100 and +100. Having agreed to put the upper limit at 1, it doesn't make sense to assume that an agent believes something to degree 7.

Axioms (ii) and (iii) look more substantive. It seems that we can at least imagine an agent who assigns degree of belief less than 1 to a logically necessary proposition, or whose credence in a disjunction of incompatible propositions is not the sum of her credence in the disjuncts. Still, we need to clarify what exactly it is that we're imagining if we want to discuss whether the imagined states are rational or irrational.

For example, suppose we understand strength of belief as a certain introspectible quantity: a special feeling of conviction people have when entertaining propositions. On this approach, axiom (ii) says that when agents entertain logically necessary propositions, they ought to experience the relevant sensation with maximal intensity. It is hard to see why this should be norm of rationality. It is also hard to see why the sensation should guide an agent's choices in line with the MEU Principle, or why it should be sensitive to the agent's evidence. In short, if we understand degrees of belief as measuring the intensity of a certain feeling, then the norms of Bayesian decision theory and Bayesian epistemology become implausible and inexplicable.

A more promising line of thought assumes that strength of belief is defined, perhaps in part, by the MEU Principle. On this approach, what we mean when we say that an agent has such-and-such degrees of belief is (in part) that she is (or ought to be) disposed to make certain choices. We can then assess the rationality of the agent's beliefs by looking at the corresponding choice dispositions.

Of course, beliefs alone do not settle choices. The agent's desires or goals also play a role. The argument we are going to look at next therefore fixes an agent's goals, by assuming that utility equals monetary payoff. Afterwards we will consider how this assumption could be relaxed.

3.2 The betting interpretation

It is instructive to compare degrees of belief with numerical quantities in other parts of science. Take mass. What do we mean when we say that an object – a chunk of iron perhaps – has a mass of 2 kg? There are no little numbers written in chunks of iron, just as there are no little numbers written in the head. As with degrees of belief, there is an element of conventionality in the way we represent masses by numbers: instead of representing the chunk's mass by the number 2, we could just as well have used a different scale on which the mass would be 2000 or 4.40925. (Appending 'kg' to the number, as opposed to 'g' or 'lb', clarifies which convention we're using.)

I am not suggesting that mass itself is conventional. Whether a chunk of iron has a mass of 2 kg is, I believe, a completely objective, mind-independent matter. If there were no humans, the chunk would still have the same mass. What's conventional is only the representation of masses (which are not intrinsically numerical) by numbers.

The reason why we can measure mass in numbers – and the reason why we know anything at all about mass – is that things tend to behave differently depending on their mass. The greater an object's mass, the harder the object is to lift up or accelerate. Numerical measures of mass reflect these dispositions, and can be standardized by reference to particular manifestations. For example, if we put two objects on opposite ends of a balance, the object with greater mass will go down. We could now choose a random chunk of iron, call it the "standard kilogram", and stipulate that something has a mass of n kg just in case it balances against n copies of the standard kilogram (or against n objects each of which balances against the standard kilogram).

Can we take a similar approach to degrees of belief? The idea would be to find a characteristic way in which degrees of belief manifest themselves in behaviour and use that to define a numerical scale for degrees of belief.

So how do you measure someone's degrees of belief? The classical answer is: by offering them a bet. Consider a bet that pays £1 if it will rain at noon tomorrow, and nothing if it won't rain. How much would you be willing to pay for this bet?

We can calculate the expected payoff – the average of the possible payoffs, weighted by their subjective probability. Let x be your degree of belief that it will rain tomorrow, and $1-x$ your degree of belief that it won't rain. The bet gives you £1 with probability x and £0 with probability $1-x$. The expected payoff is $x \cdot £1 + (1-x) \cdot £0 = £x$. This suggests that the bet is worth £x, that £x is the most you should pay for the bet.

Exercise 3.1 †

Suppose your degree of belief in rain is 0.8 (and your degree of belief in not-rain 0.2). For a price of £0.70 you can buy a bet that pays £1 if it rains and £0 if it doesn't rain. Draw a decision matrix for your decision problem and compute the expected utility of the acts, assuming your subjective utilities equal the net amount of money you have gained in the end.

If we're looking for a way to measure your degrees of belief, we can turn this line of reasoning around: if £x is the most you're willing to pay for the bet, then x is your degree of belief in the proposition that it will rain. This leads to the following suggestion.

The betting interpretation

An agent believes a proposition A to degree x just in case she would pay up to £x for a bet that pays £1 if A is true and £0 otherwise.

The betting interpretation is meant to have the same status as the above (hypothetical) stipulation that an object has a mass of n kg just in case it balances against n copies of the standard kilogram. On the betting interpretation, offering people bets is like putting objects on a balance scale. For some prices, the test person will prefer to buy the bet, for others she will prefer to sell the bet; in between there is a point at which the price of the bet is in balance with the expected payoff, so the test person will be indifferent between buying, selling, and doing neither. The price at the point of balance reveals the subject's degree of belief. The stake of £1 is a unit of measurement, much like the standard kilogram in the measurement of mass.

The betting interpretation gives us a clear grip on what it means to believe a proposition to a particular degree. It also points towards an argument for probabilism. For we can show that if an agent's degrees of belief do not satisfy the probability axioms (for short, if her beliefs are not **probabilistic**) then she is disposed to enter bets that amount to a guaranteed loss.

3.3 The Dutch Book theorem

In betting jargon, a combination of bets is called a 'book'. A combination of bets that amounts to a guaranteed loss is called a '**Dutch book**' (no-one knows why). We are going to show that if an agent's degrees of belief violate one or more of the Kolmogorov axioms, and she values bets in accordance with their expected payoff, then she will be prepared to accept a Dutch book.

The argument is a little easier to state if we look not only at bets the agent is prepared to buy, but also at bets she is prepared to sell. Selling a bet means offering it to somebody else, in exchange for a fixed amount of money. We are going to assume that if an agent's credence in a proposition A is x, then she is prepared to sell a "unit bet on A" – a bet that pays £1 in case of A and £0 otherwise – at a price of £x or more.

Exercise 3.2 ††

Show that selling a unit bet on A for £x is equivalent to buying a unit bet on $\neg A$ for £$(1 - x)$, in the sense that the two transactions have the same net effect on the decision-maker's wealth, whether or not A is true. (This means that whenever we talk about selling a unit bet in what follows, we could equivalently have talked about buying a (different, but related) unit bet.)

Now, suppose an agent's degrees of belief violate Kolmogorov's axiom (i). Concretely, suppose her credence in some proposition A is 2. By the betting interpretation, she is willing to pay up to £2 for a deal that pays her back either £0 or £1, depending on whether A is true. She is guaranteed to lose at least £1. More generally, if an agent's degree of belief in A is greater than 1, then she will be prepared to buy a unit bet on A for more than £1, which leads to a guaranteed loss.

Similarly, suppose an agent's credence in A is below 0. Let's say it is -1. The agent will then be prepared to sell a unit bet on A for any price above £-1. What does it mean to sell a bet for £-1? It means to pay someone £1 to take the bet. So the agent would pay up to £1 for us to take the bet. Having sold the bet, she will have to pay us an additional £1 if A is true. Her net loss is either £2 or £1, and guaranteed be at least £1. Again, the argument generalizes to any degree of belief below 0.

I leave the case of axiom (ii) as an exercise.

Exercise 3.3 ††

Show that if an agent's degrees of belief violate Kolmogorov's axiom (ii) then (assuming the betting interpretation) they are prepared to buy or sell bets that amount to a guaranteed loss.

Turning to axiom (iii), suppose an agent's credence in the disjunction $A \vee B$ of two logically incompatible propositions A and B is not the sum of her credence in the individual propositions. For concreteness, suppose $\text{Cr}(A) = 0.4$, $\text{Cr}(B) = 0.2$, and $\text{Cr}(A \vee B) = 0.5$. By the betting interpretation, the agent is willing to sell a unit bet on $A \vee B$ for at least £0.50. She is also willing to buy a unit bet on A for up to £0.40, and she is willing to buy a unit bet on B for up to £0.20. Notice that if she buys both of these latter bets then she has in effect bought a unit bet on $A \vee B$, for she will get £1 if either A or B is true, and £0 otherwise. So the agent is, in effect, willing to buy this bet for £0.60 and sell it for £0.50. You can check that no matter whether A or B or neither of them is true, the agent is guaranteed to lose £0.10.

The reasoning generalizes to any other case where $\text{Cr}(A \vee B)$ is less than $\text{Cr}(A) + \text{Cr}(B)$. For cases where $\text{Cr}(A \vee B)$ is greater than $\text{Cr}(A) + \text{Cr}(B)$, simply swap all occurrences of 'buy' and 'sell' in the previous paragraph.

We have proved the *Dutch Book Theorem*.

Dutch Book Theorem

Assuming the betting interpretation, any agent whose degrees of belief don't conform to the Kolmogorov axioms is prepared to buy bets whose net effect is a guaranteed loss.

One can also show the converse, that any agent who is prepared to accept a (certain kind of) Dutch book has non-probabilistic beliefs. In other words, agents whose beliefs conform to the rules of probability are *not* prepared to accept (certain kinds of) bets that amount to a guaranteed loss. This result is known as a **Converse Dutch Book Theorem**.

To prove an interesting Converse Dutch Book result, we should extend the betting interpretation so that it doesn't just cover unit bets. (We don't just want to show that an agent with probabilistic beliefs is not prepared to accept a Dutch Book *made entirely of unit bets*.) Let's assume that agents generally value bets by their expected monetary payoff, so that they pay up to £x for a bet with expected payoff £x, where the expected payoff is computed with the agent's credence function. We're now interested in cases where this credence function is a genuine probability measure, so that the expected payoff is a genuine "expectation", in the mathematical sense: a probability-weighted average.

Now consider an agent with probabilistic beliefs. If the agent pays some amount £x for a bet with expected payoff £y, then the entire transaction (including the purchase price) has expected payoff $£(y–x)$. By our extended betting interpretation, the agent makes the transaction only if $x \leq y$, in which case $£(y–x) \geq £0$. In other words, the agent makes the transaction only if the transaction has a non-negative expected payoff. Evidently, a transaction can't have a non-negative *expected* payoff unless there is at least some possibility for it to have a non-negative payoff. This shows that an agent with probabilist credences can't be "Dutch booked" with a single bet. What about combinations of bets? Suppose our agent buys a number of bets. We know that each of these transactions on its own has a non-negative expected payoff. We also know that the total payoff from all transactions together is the sum of the payoffs of the individual transactions. Now here is a useful fact about mathematical expectation: *the expectation of a sum (of some quantities) is the sum of the expectations (of the quantities)*. Since the sum of non-negative values can't be negative, this tells us that the expected total payoff from our agent's transactions isn't negative. As before, we can infer that the combined transactions are not guaranteed to generate a loss.

Exercise 3.4 ††

Here I have twice appealed to the fact that if a transaction or combination of transactions has non-negative *expected* payoff, then there must be at least a possibility of an *actual* non-negative payoff. Can you explain why this is the case? Does it depend on whether the expected payoff is computed with a genuine probability function?

Exercise 3.5 ††

Suppose I believe that it is raining to degree 0.6 and that it is not raining also to degree 0.6. Describe a Dutch book you could make against me, assuming the betting interpretation.

3.4 Problems with the betting interpretation

The Dutch Book Theorem is a mathematical result. It does not show that rational degrees of belief satisfy the probability axioms. To reach that conclusion, and thereby an argument for probabilism, we need to add some philosophical premises about rational belief.

A flat-footed "Dutch book argument" might go as follows. If your beliefs violate the probability axioms, then a cunning Dutchman might come along and trick you out of money. If your beliefs are probabilistic, he can't do that. To be safe against the Dutchman, it is better to have probabilistic beliefs.

Is this a good argument for probabilism? Two problems stand out.

First, why should the possibility of financial loss be a sign of irrational beliefs? True, there might be a Dutchman going around exploiting people with non-probabilistic beliefs. But there might also be someone (a Frenchman, say) going around richly rewarding people with non-probabilistic beliefs. We don't think the latter possibility shows that people ought to have non-probabilistic beliefs. If there is such a Frenchman, we can at most conclude that it would be *practically useful* to have non-probabilistic beliefs. But those beliefs would still be *epistemically irrational*. (Compare: if someone offers you a million pounds if you believe that the moon is made of cheese, then that belief would be practically useful, but it would not be epistemically rational.) Why should we think differently about the hypothetical Dutchman?

Second, the threat of financial exploitation only awaits non-probabilistic agents who value bets by their expected monetary payoff, as implied by the betting interpretation. Real people don't actually do this.

Consider the following gamble.

Example 3.1 (The St. Petersburg Paradox)

I am going to toss a fair coin until it lands tails. If I get tails on the first toss, I'll give you £2. If I get heads on the first toss and tails on the second, I'll give you £4. If I get heads on the first two tosses and tails on the third, I'll give you £8. In general, if the coin first lands tails on the nth toss, I'll give you £2^n.

How much would you pay for this gamble?

We can compute the expected payoff. With probability ½ you'll get £2; with probability ¼ you get £4; with probability ⅛ you get £8; and so on. The expected payoff is

$$1/2 \cdot £2 + 1/4 \cdot £4 + 1/8 \cdot £8 + \ldots = £1 + £1 + £1 + \ldots .$$

The sum of this series is infinite. If you value bets by their expected monetary payoff, you should sacrifice everything you have for an opportunity to play the gamble. In reality, few people would do that, seeing as the payoff is almost certain to be quite low.

Exercise 3.6 †

What is the probability that you will get £16 or less when playing the St. Petersburg gamble?

The St. Petersburg Paradox was first described by the Swiss mathematician Nicolas Bernoulli in 1713. It prompted his cousin Daniel Bernoulli to introduce the theoretical concept of utility as distinct from monetary payoff. As (Daniel) Bernoulli realised, "a gain of one thousand ducats is more significant to the pauper than to a rich man though both gain the same amount". In other words, most people don't regard having two million pounds as twice as good as having one million pounds: the first million would make a much greater difference to our lives than the second.

In economics terminology, what Bernoulli realised is that money has **declining marginal utility**. The 'marginal utility' of a good for an agent measures how much the agent desires a small extra amount of the good. That the marginal utility of money is declining means that the more money you have, the less you value an additional pound (or dollar or ducat).

Bernoulli had a more concrete proposal. He suggested that n units of money provide not n but $\log(n)$ units of utility. This implies that doubling your wealth always provides the same boost in utility, whether it leads from £1000 to £2000 or from £1 million to £2 million, even though the second change is much larger in absolute terms. On Bernoulli's model, the expected utility of the St. Petersburg gamble for a person with a wealth of £1000 is equivalent to the utility of getting £10.95. That's the most the agent should be willing to pay for the gamble.

Exercise 3.7 †

Suppose Bernoulli is right that owning £n has a utility of $\log(n)$. You currently have £1. For a price of £0.40 you are offered a bet that pays £1 if it will rain tomorrow (and £0 otherwise). Your degree of belief in rain tomorrow is $1/2$. Should you accept the bet? Draw the decision matrix and compute the expected utilities. (You need to know that $\log(1) = 0$, $\log(1.6) \approx 0.47$, and

$\log(0.6) \approx -0.51$. Apart from that you don't need to know what 'log' means.)

Exercise 3.8 ††

As Bernoulli noticed, the declining marginal utility of money can explain the usefulness of insurance. Suppose your net worth is £100 000, and there's a 5% chance of a catastrophic event that would leave you with only £10 000. For a fee of £10 000, a bank offers you an insurance against the catastrophic event that pays £80 000 if the event occurs (and nothing otherwise). Explain (informally, if you want) why this might be a good deal both for you and for the bank.

Exercise 3.9 †

Bernoulli's logarithmic model is obviously a simplification. Suppose you want to take a bus home. The fare is £1.70 but you only have £1.50. If you can't take the bus, you'll have to walk for 50 minutes through the rain. A stranger at the bus stop offers you a deal: if you give her your £1.50, she will toss a coin and pay you back £1.70 on heads or £0 on tails. Explain (briefly and informally) why it would be rational for you to accept the offer.

There's a second reason why rational agents wouldn't always value bets by their expected payoff even if their subjective utility were adequately measured by monetary payoff. The reason is that buying or selling bets can alter the relevant beliefs.

For example, I am quite confident I will not buy any bets today. Should I therefore be prepared to pay close to £1 for a unit bet that I don't buy any bets today? Clearly not. By buying the bet, I would render the proposition false. Given my current state of belief, the (imaginary) bet has an expected payoff close to £1, but it would be irrational for me to buy it even for £0.10.

In sum, we can't assume that rational agents always value bets by their expected payoff. The betting interpretation is indefensible.

This is a setback on two fronts. One, we have lost an attractive answer to how degrees of belief are measured or defined. If an agent's degrees of belief aren't defined by their betting behaviour, then how *are* they defined? Second, and relatedly, we have lost what looked like an attractive argument for probabilism. If agents don't value bets by their monetary payoff, we can't show that non-probabilistic agents will be prepared to buy bets that amount to a sure loss.

We will look at alternative approaches to measuring belief in sections 3.6 and 6.5. First, let me explain how we might rescue an argument for probabilism from the wreckage of the betting interpretation.

3.5 A Dutch book argument

We want to show that non-probabilistic beliefs are irrational. Let α be an arbitrary agent with non-probabilistic beliefs. We can't assume that α values bets by their expected monetary payoff. But let's imagine a counterpart β of α who has the exact same beliefs as α but possibly different, and somewhat peculiar desires. β's only goal is to increase her wealth. Money does not have declining marginal utility for β. She would give all she has for an opportunity to play the St. Petersburg gamble. β might also differ from α in another respect: whenever she faces a choice, β chooses an option that maximizes expected utility.

I'm going to need a number of philosophical assumptions. Here is the first: *if α's belief state is epistemically rational, then so is β's.* The idea is that if you want to know if someone's beliefs are epistemically rational (rather than, say, practically useful), then you need to know what her beliefs are and maybe how she acquired those beliefs, but you don't need to know what she desires or how she chooses between available acts.

As we saw at the end of the previous section, we can't assume that β will always pay up to £Cr(A) for a unit bet on A (where Cr is her credence function), since her credence in A may be affected by the transaction. But this problem only seems to arise for a small and special class of propositions. Let's call a proposition *stable* if it is probabilistically independent, in β's credence function, of the assumption that she buys or sells any particular bets. The probability axioms are supposed to be general consistency requirements on rational belief. Such requirements should plausibly be "topic-neutral": they should hold for beliefs or every kind, not just for beliefs about a special subject matter. In particular, there aren't special consistency requirements that only pertain to stable beliefs. *If an agent's credences over stable propositions should be probabilistic, then her entire credence function should be probabilistic.*

This is my second assumption. It implies that in order to show that non-probabilistic beliefs are irrational, it suffices to show that non-probabilistic beliefs towards stable propositions are irrational. So we can assume without loss of generality that α's (and therefore β's) beliefs towards stable propositions are non-probabilistic.

We know that if a proposition A is stable, then β is prepared to pay up to £$\mathrm{Cr}(A)$ for a unit bet on A. That's because β's utility function simply measures monetary payoff and because she obeys the MEU Principle. The betting interpretation is correct for β, as long as we stick with stable propositions.

We also know that β's credences towards stable propositions violate the probability axioms. It follows by the Dutch Book Theorem that she is prepared to buy bets whose net effect is a guaranteed loss. My next assumption states that it would be irrational for β to make these transactions: *it is irrational for an agent whose sole aim is to increase her wealth to (deliberately and avoidably) make choices whose net effect is a guaranteed loss*.

This was my third assumption. My fourth assumption is that irrational choices always arise from either irrational beliefs or from irrational desires or from an irrational way of linking up one's beliefs and desires to one's actions. I also assume that the right way of linking up beliefs and desires to actions is given by the MEU Principle. Thus: *if an agent is disposed to make irrational choices, then she is either epistemically irrational, or her desires are irrational, or her acts don't maximize expected utility*.

In the case of β, we can rule out the third possibility. Her choices do maximize expected utility. I also claim (assumption 5) that *β's desires are not irrational*. Admittedly, her desires are odd. We might call them unreasonable, or even "irrational" in a substantive sense. But they aren't inconsistent. They represent a coherent evaluative perspective.

Since β is disposed to make irrational choices, we can infer that she is epistemically irrational. By the very first assumption, it follows that α is epistemically irrational. And α was an arbitrary agent whose credences violate the rules of probability. We've shown that (epistemically) rational beliefs are probabilistic.

My argument relies on a lot of assumptions. Many of them could be challenged. Can you think of a better argument?

3.6 Comparative credence

We have seen that the betting interpretation is untenable. Many philosophers hold that degrees of belief cannot be defined in terms of an agent's behaviour, but should rather be treated as theoretical primitives. Even on that view, however, more must be said about the numerical representation of credence. That we represent degrees of belief by numbers between 0 and 1 is clearly a matter of convention. We need to explain how this convention of assigning numbers to propositions works.

One approach towards such an explanation, which does not turn on an agent's behaviour, was outlined by the Italian mathematician and philosopher Bruno de Finetti (who, incidentally, also published the first proof of the Dutch Book Theorem). De Finetti suggested that degrees of belief might be defined in terms of the comparative attitude of being more confident in one proposition than in another. While any numerical representation of beliefs is partly conventional, this comparative attitude is plausibly objective and might be taken as primitive.

Let '$A \succ B$' express that a particular (not further specified) agent is more confident in A than in B. For example, if you are more confident that it is sunny than that it is raining, then we have $Sunny \succ Rainy$. Let '$A \sim B$' mean that the agent is equally confident in A and in B. From these, we can define a third relation '$\succsim$' by stipulating that $A \succsim B$ iff $A \succ B$ or $A \sim B$.

We now make some assumptions about the formal structure of these relations. To begin, if you are more confident in A than in B, then you can't also be more confident in B than in A, or equally confident in the two. We also assume that if you're neither more confident in A than in B, nor in B than in A, then you're equally confident in A and B. Your comparative credence relations are then "complete", in the following sense:

Completeness

For any A and B, exactly one of $A \succ B$, $B \succ A$, or $A \sim B$ is the case.

Next, suppose you are at least as confident in A as in B, and at least as confident in B as in C. Then you should be at least as confident in A as in C. So $\succsim$ should be "transitive":

Transitivity

If $A \succsim B$ and $B \succsim C$ then $A \succsim C$.

Exercise 3.10 †††

Show that Transitivity and Completeness together entail that (a) if $A \sim B$ then $B \sim A$, and (b) if $A \sim B$ and $B \sim C$, then $A \sim C$.

For the next assumptions, I use '$\top$' to stand for the logically necessary proposition (the set of all worlds) and '$\bot$' for the logically impossible proposition (the empty set).

Non-Trviality

$\top \succ \bot$.

Boundedness

There is no proposition A such that $\bot \succ A$.

These should be fairly plausible demands of rationality.

My next assumption is best introduced by an example. Suppose you are more confident that Bob is German than that he is French. Then you should also be more confident that Bob is *either German or Russian* than that he is *either French or Russian*. Conversely, if you are more confident that he is German or Russian than that he is French or Russian, then you should be more confident that he is German than that he is French. In general:

Quasi-Additivity

If A and B are both logically incompatible with C, then $A \succsim B$ iff $(A \vee C) \succsim (B \vee C)$.

De Finetti conjectured that whenever an agent's comparative credence relations satisfy the above five assumptions, then there is a unique probability measure Cr such that $A \succsim B$ iff $\mathrm{Cr}(A) \geq \mathrm{Cr}(B)$ (which entails that $A \succ B$ iff $\mathrm{Cr}(A) > \mathrm{Cr}(B)$ and $A \sim B$ iff $\mathrm{Cr}(A) = \mathrm{Cr}(B)$). The conjecture turned out to be false, because a sixth assumption is required. But the following can be shown:

Probability Representation Theorem

If an agent's comparative credence relations satisfy Completeness, Transitivity, Non-Triviality, Boundedness, Quasi-Additivity, and the Sixth Assumption, then there is a unique probability measure Cr such that $A \succsim B$ iff $\mathrm{Cr}(A) \geq \mathrm{Cr}(B)$.

Before I describe the Sixth Assumption, let me explain what the Probability Representation Theorem might do for us.

I have argued that we can't take numerical credences as unanalysed primitives. There must be an answer to why an agent's degree of belief in rain is correctly represented by the number 0.2 rather than, say, 0.3. De Finetti's idea was to derive numerical representations of belief from comparative attitudes towards propositions.

Imagine we order all propositions on a line, in accordance with the agent's comparative judgements (which we take as primitive). Whenever the agent is more confident in one proposition than in another, the first goes to the right of the first. Whenever the agent is equally confident in two propositions, they are stacked on top of each other at the same point on the line. If the agent is reasonable, the impossible proposition $\bot$ will be at the left end, the necessary proposition $\top$ at the right end.

We now want to use numbers to represent the relative position of propositions along the line, in such a way that as we move from the $\bot$ position to the $\top$ position, the numbers get higher and higher. The Probability Representation Theorem assures us that this can be done, provided that the agent's comparative judgements satisfy the six assumptions. In that case, it says, there will be an assignment of numbers to propositions that "represents" the agent's comparative judgements in the sense that $A \succsim B$ iff the number assigned to A is at least as great as the number assigned to B.

The next problem is that if there is one such assignment then there are infinitely many, giving different numbers to propositions in between $\bot$ and $\top$. (For example, if f represents $\succsim$ then so does the function g defined by $g(A) = f(A)^2$.) We need to settle on a particular assignment. Again, the Probability Representation Theorem comes to our help. It tells us that among the eligible assignments of numbers to propositions – among those that represent the agent's comparative judgements – there is only one that satisfies the conditions on a probability measure. Let's adopt the convention of using this assignment.

On this approach, '$\text{Cr}(Rain) = 0.2$' means that the agent's comparative confidence judgements order the propositions in such a way that the unique probability measure that "represents" these judgements assigns 0.2 to *Rain*. Any agent whose attitudes of comparative credence satisfy the six assumptions is guaranteed to have probabilistic credences, because the agent's credence function is *defined* as the unique probability measure (!) that represents her comparative judgements. An agent who doesn't satisfy the six assumptions doesn't have a credence function at all, because our convention of measurement – on the present approach – doesn't cover such agents.

As you may imagine, this approach has also not gone unchallenged. One obvious question is whether we can take comparative confidence as primitive. If we can, a further question is whether the six assumptions are plausible as general constraints on any agent with degrees of belief. The missing sixth assumption is especially troublesome in this regard. The form of the assumption turns out to depend on whether the number of propositions is finite or infinite. In either case the condition is so complicated that many struggle to accept it as a basic norm of rationality – let alone as a basic condition anyone must satisfy in order to have degrees of belief at all. Just to prove the point, here is the condition for the slightly simpler case of finitely many propositions:

The Sixth Assumption (finite version)

For any two sequences of propositions $A_1, \dots, A_n$ and $B_1, \dots, B_n$ such that for every possible world w there are equally many propositions in the first sequence that contain w as in the second, if $A_i \succsim B_i$ for all $i < n$, then $B_n \succsim A_n$.

Essay Question 3.1

I have expressed the Dutch Book Theorem with monetary outcomes. One might try to avoid commitment to the betting interpretation by replacing the monetary outcomes with other goods the agent happens to care about. For example, when we looked at Kolmogorov's axiom (i), I said that an agent whose degree of belief in A is 2 would pay (say) £1.50 for a bet that pays £1 if A is true and £0 otherwise. This assumes the betting interpretation. Now let 'U1.5' denote an arbitrary good to which the agent assigns utility 1.5. Similarly, let U1 be a good with utility 1, and U0 a good with utility 0. Consider a bet that would give the agent U1 if A is true and U0 otherwise. The bet's expected utility is $\mathrm{Cr}(A) \cdot \mathrm{U}(\mathrm{U1}) + \mathrm{Cr}(\neg A) \cdot \mathrm{U}(\mathrm{U0}) = \mathrm{Cr}(A)$. Assuming the MEU Principle, an agent with $\mathrm{Cr}(A) = 2$ would prefer this bet over U1.5, even though the latter is guaranteed to give her greater utility, which is surely irrational. Can you spell out a full argument for probabilism along these lines? What problems do you see for this line of argument?

Sources and Further Reading

For a critical overview and assessment of Dutch Book arguments, see Alan Hájek, "Dutch Book Arguments" (2008). If you want to dive even deeper, you may start with Susan Vineberg's Stanford Encyclopedia entry on Dutch Book Arguments (2022).

For a more extensive philosophical introduction and criticism of the comparative approach from section 3.6, see Edward Elliott, "Comparativism and the Measurement of Partial Belief" (2020). Peter Fishburn's "The Axioms of Subjective Probability" (1986) goes deeper into the mathematical background.

A recently popular third way of arguing for probabilism, besides the Dutch Book approach and the comparative approach, draws on the observation (also first made by de Finetti) that for every non-probabilistic credence function there is a probabilistic credence function that is guaranteed to be closer to the truth – where closeness to the truth is a certain measure of the distance between the credence given to any proposition and the proposition's truth-value (0=false, 1=true). See, for example, James Joyce, "A nonpragmatic vindication of probabilism" (1998).

Martin Peterson's Stanford Encyclopedia entry on the St. Petersburg paradox discusses the historical context of the St. Petersburg paradox and also introduces a "modern" version in which the monetary payoffs are replaced by units of utility.

The bus fare exercise is from Brian Skyrms, *Choice and Chance* (2000).