Ramsey's Representation Theorem

Richard BRADLEY†

ABSTRACT

This paper reconstructs and evaluates the representation theorem presented by Ramsey in his essay 'Truth and Probability', showing how its proof depends on a novel application of Hölder's theory of measurement. I argue that it must be understood as a solution to the problem of measuring partial belief, a solution that in many ways remains unsurpassed. Finally I show that the method it employs may be interpreted in such a way as to avoid a well known objection to it due to Richard Jeffrey.

1. The Measurement Problem

Ramsey's essay *'Truth and Probability*' and, in particular, the representation theorem that he sketches in the third section of it, is widely regarded as having anticipated subsequent work in decision theory and the theory of subjective probability. Both Leonard Savage and Bruno De Finetti, for instance, attest to the significance of Ramsey's work, while others, such as Ethan Bolker (1967), Richard Jeffrey (1983) and Donald Davidson and Patrick Suppes (1956) acknowledge its influence on their own. Despite this, Ramsey's representation theorem remains by and large poorly understood, both with regard to how the theorem works and what he is trying to achieve by it. One reason is that there has been little recognition in decision theory of the distinction between the problem of justifying its theoretical claims regarding the properties of rational belief and desire and the problem of measuring its variables – degrees of belief and desire. But Ramsey must also take some responsibility for this situation, for he failed to work out all the details of his theory, claiming that:

> this would, I think, be rather like working out to seven places of decimals a result only valid to two (Ramsey 1926, 76)

† Department of Philosophy, Logic and Scientific Method, London School of Economics, Email: r.bradley@lse.ac.uk

The result has been an unfortunate neglect of some very important features Ramsey's work; features that surpass rather than anticipate subsequent work in decision theory.

1.1 Ramsey's Problem

The main focus of 'Truth and Probability' is the development of the idea that probability theory gives a logic of partial belief. Part 3 of the essay is devoted to identifying the relevant concept of partial belief – namely that of basis or causal component of action – and to the related question of how it might be measured. Ramsey evidently felt that the concept of partial belief would lack definiteness unless one could say how, in principle if not in practice, this might be done.

> the degree of belief is like a time interval; it has no precise meaning unless we specify how it is to be measured (*ibid*, 63)

But even though most of his attention is directed at this measurement problem, Ramsey was no operationalist: as we shall see later on, he clearly recognised the impossibility of specifying a procedure for measuring partial belief (or any other quantity) without a rather rich conception of what is being measured and of what properties it could be expected to display.

Ramsey's thinking on the problem of measuring belief seems at first to be very much in the behaviourist mould. He argues, for instance, that the idea that believing something more or less strongly was connected to a perceptible feeling of belief of a certain intensity is:

> ... observably false, for the beliefs which we hold most strongly are often accompanied by practically no feeling at all; no one feels strongly about things he takes for granted. (*ibid*, 65)

And that:

> when we seek to know what is the difference between believing more firmly and believing less firmly, we can no longer regard it as consisting in having more or less observable feelings; at least I personally cannot recognise any such feelings. The difference seems to me to lie in how far we should act on these beliefs ... (*ibid*, 66)

But whereas the behaviourist rejects introspective evidence of any kind on the grounds that it is not intersubjectively verifiable, Ramsey rejects only the particular use of introspection associated with the idea of measuring strength of belief in terms of the sensations or feelings that accompany it. Indeed in the argument just quoted he makes use of introspection in citing his own failure to perceive a feeling corresponding to his belief. And further on, when he argues that although we may feel that:

> we know how strongly we believe things and that we can only know this if we can measure our belief by introspection ... our judgement about the strength of our belief is really about how we should act in hypothetical circumstances. (*ibid*, 67)

the judgement that he refers to – as to how we would act under hypothetical circumstances – is presumably an introspective one.

Ramsey takes his arguments to show that we can although we might be able to introspect whether we do or do not believe something, there is no reliable way of introspecting the degree to which we do. It would appear that this suspicion of introspection, if not his arguments against it, extends to the possibility of qualitative judgements as to whether one believes one thing more strongly than another, despite their apparent similarity to judgements as to how we would choose or act in particular circumstances. The upshot is that he admits only evidence as to the choices that an agent does or would make between specified alternatives and not the agent's direct reports on their partial attitudes.

1.2 Ramsey's Solution

Ramsey's problem may now be reformulated as that of explaining how the evidence relating to an agent's choices, or rather the preferences implied by them, determines a measure of her degrees of partial belief. Now any approach to the measurement of belief that admits only evidence of this kind faces what is frequently termed the problem of the simultaneous determination of belief and desire: essentially that of untangling the respective contributions made by an agent's beliefs and desires to her choices. Ramsey's solution in a nutshell is as follows. First he introduces the concept of what I will call a conditional prospect. Conditional prospects are the sorts of possibilities we ordinarily express by the use of (indicative) conditional sentences, such as that if it's hot today, then it will rain tomorrow and, if not, it will snow.[1] Next Ramsey shows how evidence pertaining to an agent's preferences amongst conditional prospects allows us to identify ethically neutral propositions – propositions whose truth or otherwise are of no consequence to the agent – that the agent believes are as likely to be true as not. Then he uses these propositions to measure the agent's degrees of preference for the various prospects under consideration. Finally, the agent's degrees of partial belief are derived from her degrees of preference.

Somewhat more formally, what Ramsey gives us is a method of constructing a pair of real-numbered functions (let us call them *bel* and *des*) that re-

[1] Such prospects have come to be termed 'gambles' or 'actions' in the literature on Ramsey, although he does not use either term. Both terms are misleading.

spectively assign to each possible object of the agent's attitude a measure of the strength of the agent's belief in, and desire for, the possibility it expresses. A crucial feature of his method is that it presupposes the theory that we act or choose in a way which has the best expected consequences, given our partial beliefs, a theory which he claims

> cannot be made adequate to all the facts, but ... [is] a useful approximation to the truth, particularly in our self-conscious or professional life, and it is presupposed in a great deal of our thought. (*ibid,* 69)

Measurement never takes place in a theoretical vacuum, of course, and it would be quite wrong to accuse Ramsey of simply presupposing what he needs to show. For the task Ramsey gives himself is not that of justifying the particular theory of rational action that he adopts but of showing how, given this theory, it is possible to measure degrees of belief. But it is important to recognise that Ramsey has to take on board much more than in contained in the folk-psychological idea that action is guided by expected benefit. For his method assumes, in a way that will become clear later on, that the desirability of a conditional prospect is related in a precise way to the desirability of the possible states of the world consistent with it; namely, that the former is a weighted average of the latter, with the weights coming from the agent's degrees of belief. Formally, the desirability of the conditional prospect γ, that X be the case if P is and that Y be the case if not, is assumed to be given by the expression:

Proposition 1: $des(\gamma) = des(X \text{ and } P) \,.\, bel(P) + des(Y \text{ and } \neg P)\,(1 - bel(P))$

This proposition expresses much of the content of what we now call decision theory, but Ramsey offers little in the way of a justification for it. Some is ultimately to be drawn from the plausibility of the axioms of his representation theorem, but these seem not to be have formulated with justification in mind. Given that his claim that degrees of belief are probabilities is established by his method for measuring belief and that his method depends on the expected benefit assumption, the failure to argue for this claim clearly weakens his argument that probability theory is a logic of partial belief. But, as I will argue, this hardly diminishes the interest of his method to contemporary decision theory.

2. A Method for Measuring Belief

Ramsey did not give a full explicit account of his method and our reconstruction of it must draw from what is implicit in his representation theorem. In doing so we will deviate from his account in one important respect. Ramsey makes a distinction between the objects of belief – propositions – and the objects of desire – prospects. Prospects are of two types: possible courses of the world (worlds for short) and conditional prospects, which are essentially functions from partitions of propositions to worlds. But Ramsey's worlds are most satisfactorily understood to be propositions that are maximally specific with regard to matters of concern to the agent, so we can simplify his ontology and work just with propositions and conditional prospects.

Basic Notation

We denote propositions by upper case Roman letters with $\neg$, $\vee$, and $\wedge$ denoting the negation, disjunction and conjunction operations respectively. Arbitrary prospects (conditional or otherwise) are denoted by Greek letters. Conditional prospects are identified by expressions of the form $(\alpha \text{ if } P)(\beta \text{ if } \neg P)$. We take for granted that $(\alpha \text{ if } P)(\beta \text{ if } \neg P)$ is, for instance, the same conditional prospect as $(\beta \text{ if } \neg P)(\alpha \text{ if } P)$. Finally we denote the agent's preference for the prospect over the prospect ψ by $\varphi > \psi$ and her indifference between the two by $\varphi \approx \psi$. By $\varphi \geq \psi$ is meant that $\varphi > \psi$ or $\varphi \approx \psi$.

Step 1: Defining Ethical Neutrality

Ramsey's critical innovation is the postulation of what he calls ethically neutral propositions. An ethically neutral proposition is simply one whose truth or falsity is a matter of indifference to the agent and does not affect their attitude to any other prospects e.g. the prospect of a dust storm on Mars does not influence any of my earthly concerns. Crucially the probabilities of some ethically neutral propositions can be inferred from an agent's choices. Suppose, for instance, that an agent is not indifferent between the prospect of sun and that of snow, but indifferent between the prospect that if P is true, then it will be sunny, but if P is not, then it will snow, and the prospect that if P is true, then it will snow, but if P is not, then it will be sunny. Then we can infer that they regard P as likely to be true as not. For were it not, they should prefer one of the conditional prospects over the other (this follows directly from Proposition 1).

Are there any ethically neutral propositions? The standard candidates are propositions such as that the next card drawn will be an Ace or the coin will land heads. But the truth of such propositions do affect agents' attitudes to some prospects. Take any proposition X and suppose that the agent is not indifferent to prospect that A. Then the truth of X will be a matter of consequence to her attitude to the (conditional) prospect that A if X, because in the event that X is true, the prospect that A if X amounts to that of A and X. So X is not ethically neutral. It follows that there are no propositions that are neutral with respect to all prospects. On the other hand, we really only need to assume the existence of propositions whose truth or falsity is of no consequence to the agents' attitudes to other *propositions*. This idea is captured by the following:

Definition: A proposition P is ethically neutral iff for all propositions Q, $PQ \approx Q \approx \neg PQ$.

The most crucial step in the argument can now be taken. For notice that it follows from Proposition 1 that if P is ethically neutral then the desirability of the prospect (X if P) (Y if $\neg$P) equals that of the prospect (Y if P) (X if $\neg$P) just in case:

$$des\,(X).\,bel(P) + des(Y).\,(1 - bel(P)) = des(Y).bel(P) + des(X).(1 - bel(P))$$

But this can only hold in case $bel(P) = 1 - bel(P)$ i.e. in case $bel(P) = 0.5$. This is the justification for the following definition of the property of having subjective probability one-half, attributable to ethically neutral propositions solely from the evidence of preferences expressed the agent.

Definition: Let P be any ethically neutral proposition and X and Y be any propositions that are consistent with the truth of both P and $\neg P$ and such that $X > Y$. Then P is said to be of *probability one-half iff:* (X *if* P)(Y *if* $\neg P$) ≈ (Y *if* P)(X *if* $\neg P$)

Step 2: Defining intervals of values

Suppose that P is an ethically neutral proposition of probability one-half. Now on the assumption that we have before us the agent preference ranking of all prospects, we can begin to construct a scale again which the desirability of these prospects can be measured. Let us call the set of prospects equally preferred to a prospect φ, its *value* and denote it by φ. Values function as a kind of intermediate object between prospects and the real numbers we wish to assign to them, qua measure of their desirability.

Next we introduce the idea of the difference between two values $\overline{\alpha}$ and $\overline{\beta}$ and denote it by $\overline{\alpha} - \overline{\beta}$. More exactly we say what it is for the difference between the values $\overline{\alpha}$ and $\overline{\beta}$ to exceed that between values $\overline{\gamma}$ and $\overline{\delta}$ by defining, as follows, a 'greater than' relation $\succeq$ on the set of all such differences in values. Suppose that P is an ethically neutral proposition of probability one-half. Then:

Definition: $\overline{\alpha} - \overline{\beta} \succeq \overline{\gamma} - \overline{\delta}$ iff (α if P)(δ if $\neg P$) $\geq$ (β if P)(γ if $\neg P$)

The motivation for this definition once again derives from Proposition 1, for it follows from this proposition and the fact that $bel(P) = 0.5$ that:

$des(\alpha \text{ if } P)(\delta{+}P) \geq des(\beta \text{ if } P)(\gamma \text{ if } \neg P)$

$$\Leftrightarrow \frac{des(\alpha) + des(\delta)}{2} = \frac{des(\beta) + des(\gamma)}{2}$$

$$\Leftrightarrow des(\alpha) + des(\beta) = des(\gamma) + des(\delta)$$

Equally we note that the relation $\succeq$ is observable in the sense that it can be constructed from the agent's preferences with the guidance of Proposition 1.

Step 3: Constructing a Desirability Scale

We now construct a desirability function *des* as follows. Let α and β be any two prospects such that $\alpha \geq \beta$. We set $des(\overline{\alpha}) = 1$ and $des(\overline{\beta}) = 0$. Let P be an ethically neutral proposition of probability one-half. Then to find the mid-point between these values we then look for the prospect γ such that $(\gamma \text{ if } P)(\gamma \text{ if } \neg P) \approx (\alpha \text{ if } P)(\beta \text{ if } \neg P)$ and assign it the number 0.5. We then find the midpoint between γ and β – the prospect ε such that $(\varepsilon \text{ if } P)(\varepsilon \text{ if } \neg P) \approx (\gamma \text{ if } P)(\beta \text{ if } \neg P)$ – and assign it the number 0.75. And so on until we have a have constructed a desirability scale of sufficient fineness to measure any value lying between $\overline{\alpha}$ and $\overline{\beta}$. To measure values lying above them we extend the scale by finding the value such that $\overline{\beta}$ is the midpoint between it and and assigning it the number 2. To measure values lying below $\overline{\alpha}$ and $\overline{\beta}$ we find the value such that $\overline{\alpha}$ is the midpoint between it and $\overline{\beta}$ and assign it the number -1. And so on until our 'ruler' extends to all prospects entertained by the agent.

Step 4: Deriving a Measure of Degrees of Belief

Ramsey's final move is to use the desirability measure to determine the agent's degrees of belief in all propositions, including those that are not ethically neutral. The vehicle for doing so is the following reorganisation of Proposition 1 allowing derivation of *bel* from *des*.

Definition: Suppose that $X \not\approx Y$ and $\zeta \approx (X \text{ if } P)(Y \text{ if } \neg P)$. Then:

$$\text{bel(P)} =_{defn} \frac{des(\zeta) - des(Y)}{des(X) - des(Y)}$$

Ramsey notes that in this definition the proposition P is not assumed to be ethically neutral, but that it is necessary to assume both that this definition is independent of the choice of prospects meeting the antecedent conditions and that

> there is a world with any assigned value in which P is true, and one in which P is false (*ibid*, 75)

The first assumption is obviously required if the definition of the measure *bel* is to be coherent, but Ramsey offers no explanation for the necessity of the second. It turns out however that it is required for his subsequent proof that *bel* is a probability measure (and in particular that it is additive). This is rather unfortunate for Ramsey, given that the proof is crucial to the defence of his subjective interpretation of the probability concept. For the assumption in question is in fact inconsistent with his framework.[2] Once worlds are dispensed with however, it becomes possible to express the relevant condition without fear of inconsistency.

Proposition: (Conditionalism) For any propositions P and Q, there exists a proposition R such that $PR \approx Q$.

Conditionalism is to my mind an imminently defensible doctrine. Essentially the conditionalist's claim is that however good (or bad) some possibility might be on average, there are imaginable circumstances in which it is not so. No prospect is good or bad in itself, but is only so relative to the conditions under which it is expected to be realised. And given conditionalism it is possible to prove that the function *bel* is a probability measure. We omit the details here since our concern is not with the viability of his interpretation of the probability calculus.[3] This completes our reconstruction of Ramsey's method.

3. Ramsey's Representation Theorem

Under what conditions will Ramsey's method work in the sense of allowing for the construction of *des* and *bel* from the agent's preferences? Ramsey's answer to this question takes the form of a representation theorem. The theorem shows that any preferences obeying a particular set of axioms can be represented by quantitative measures of the agent's degrees of desire and belief that jointly satisfy Proposition 1, in just the sense that degrees of belief and desire of this form imply preferences that obey the axioms. Ramsey does not give a full proof of his theorem and there seems to be very little recognition of the fact that the strategy for proving it that he sketches differs to a considerable extent from others to be found in the decision theoretic literature. My intention here is to reconstruct Ramsey's theorem to the extent necessary to make clear how his proof works.

While Ramsey directly axiomatises the ordering relation on values induced by the preference relation on prospects, I will state the axioms in terms of the latter. His way of doing it obscures some issues of importance to our discus-

[2] See Bradley (2001).
[3] A detailed reconstruction of this proof can be found in Bradley (2001).

sion and is easily recovered from ours. Let Δ be a non-empty set of ethically neutral propositions of probability one-half and suppose that P belongs to Δ. Then Ramsey postulates:

R1 If $Q \in \Delta$ and (α if P)(β if $\neg P$) $\geq$ (γ if P)(δ if $\neg P$), then:
(α if Q)(β if $\neg Q$) $\geq$ (γ if Q)(δ if $\neg P$)

R2 If (α if P)(δ if $\neg P$) $\approx$ (β if P)(γ if $\neg P$) then:
(i) $\alpha > \beta \Leftrightarrow \gamma > \delta$
(ii) $\alpha \approx \beta \Leftrightarrow \gamma \approx \delta$

R3 (i) $\varphi \geq \psi$ or $\psi > \varphi$
(ii) If $\varphi \geq \psi$ and $\psi \geq \omega$, then $\varphi \geq \omega$.

R4 If (α if P)(δ if $\neg P$) $\geq$ (β if P)(γ if $\neg P$) and (γ if P)(ζ if $\neg P$) $\geq$ (δ if P)(η if $\neg P$), then (α if P)(ζ if $\neg P$) $\geq$ (β if P)(η if $\neg P$)

R5 $\forall(\alpha,\beta,\gamma)[\exists(\delta)$:($\alpha$ if P)(γ if $\neg P$) $\approx$ (δ if P)(β if $\neg P$)]

R6 $\forall(\alpha,\beta)[\exists(\delta)$:($\alpha$ if P)(β if $\neg P$) $\approx$ (δ if P)(δ if $\neg P$)]

R7 Archimedean Axiom.

I have slightly strengthened Ramsey's first, third and fourth axiom by stating them in terms of the weak order $\geq$ rather than the indifference relation $\approx$. In the presence of R2, my R5 and R6 are jointly equivalent to his fifth and sixth axioms. Ramsey doesn't say what he intends by R7, but its significance, whatever its precise formulation, is that it allows the derivation of the Archimedean condition referred to in the Definition below. Ramsey also postulates one further axiom – a continuity condition – that is redundant once worlds have been dispensed with and so has been omitted here.

The role of these axioms is to ensure the existence of a numerical representation, not only of such facts as the agent preferring one thing to another, but also of the extent to which the strength of her preference for one thing exceeds the strength of her preference for another. Axiom R1 is the one that most characterises his method for it is this axiom that ensures the coherence of his definition of an ordering of differences in values. Axioms R5, R6 and R7 are there to ensure a correspondence between values and real numbers. Finally, R2, R3 and R4 ensure that the difference operation on values functions like the subtraction operation on real numbers. In particular, note that Axiom R4, opaque in its original formulation, translates as an axiom of transitivity for differences in values:

R4* If $\overline{\alpha} - \overline{\beta} \succeq \overline{\gamma} - \overline{\delta}$ and $\overline{\gamma} - \overline{\delta} \succeq \overline{\eta} - \overline{\zeta}$, then $\overline{\alpha} - \overline{\beta} \succeq \overline{\eta} - \overline{\zeta}$

We are now in a position to state Ramsey's theorem establishing the existence of a measure of the agent's desires. Ramsey does not give a uniqueness theorem for such measures, but his subsequent discussion of the measurement of probabilities assumes that they are unique up to affine linear transformation (or choice of scale). Since these uniqueness properties also follow from his axioms we fold them into our statement of his theorem.

Representation Theorem: There exists a utility function, *des*, on the set of values, such that $\forall(\bar{\alpha}, \bar{\beta}, \bar{\gamma}, \bar{\delta})$:

$$des(\bar{\alpha}) - des(\bar{\beta}) \geq des(\bar{\gamma}) - des(\bar{\delta}) \Leftrightarrow (\alpha \text{ if } P)(\beta \text{ if} \neg P) \geq (\gamma \text{ if } P)(\delta \text{ if} \neg P)$$

Furthermore, if *des*' is another such a function on the set of values, then there exists real numbers a and b, such that $a > 0$ and $des' = a.des + b$.

The key to understanding Ramsey's representation theorem is to recognise that it implicitly draws on the theory of measurement deriving from the work of the German mathematician Hölder (with which Ramsey would have been familiar). We begin with a statement of the relevant results in this area, drawing from their presentation in Krantz et al.[4]

Definition: Let A be a non-empty set and $\succeq$ a binary relation on $A \times A$. Then $< A \text{ x } A, \succeq >$ is an *algebraic difference structure* iff $\forall$ $(a,b,c,d,a'b'c' \in A)$:

1. $\succeq$ is a complete and transitive
2. If $ab \succeq cd$, then $dc \succeq ba$
3. If $ab \succeq cd$ and $bc \succeq b'c'$ then $ac \succeq a'c'$
4. If $ab \succeq cd$ then there exists $x,x' \in A$, such that $ax \approx cd \approx x'b$
5. Archimedean condition

Measurement Theorem: If $<\text{A} \times \text{A}, \succeq>$ is an algebraic difference structure, then there exists a real-valued function, φ, on A, such that $\forall(a,b,c,d \in A)$:

$$ab \succeq cd \Leftrightarrow \varphi(a) - \varphi(b) \geq \varphi(c) - \varphi(d)$$

Furthermore, φ is unique up to positive linear transformation i.e. if ψ is another such a function then $\exists x, y \in \Re : x > 0$, $\psi = x. + y.$

Ramsey's strategy for proving his representation theorem essentially consists in the use of preference orderings of prospects to define an algebraic difference structure and then to invoke the Measurement Theorem. For instance, the completeness of the ordering on differences in values follows immediately from that of $\geq$, while its transitivity is exactly what is postulated by R4. Axiom

[4] The authors point out that Hölder's results can be applied to the problem of measurement of degrees of preference, but (oddly) make no attempt do so directly. Nor is there explicit recognition of the use that Ramsey makes of them.

R4 is again essential, in conjunction with R2, to the demonstration that the second and third conditions hold, while the fifth follows from R5.

4. Ethical Neutrality versus State-Independence

> Ramsey's essays, though now much appreciated, seem to have had relatively little influence. (Savage 1954, 96)

Savage's remark applies equally well today and mainstream decision theory descends from Savage and not Ramsey. There are, I think, two reasons for his lack of influence. One is that Ramsey's style is so elliptical, and his writings so lacking in detail, that decision theorists have been unsure as to what exactly he has or has not achieved.[5] The second is that the distinction between the problem of justifying the claims of decision theory regarding the properties of rational belief and desire and the problem of the measurement of the decision theoretic variables – degrees of belief and desire – has not been properly recognised. This is largely due, I suspect, to the fact that decision theoretic representation theorems can be applied to both kinds of problems, depending on whether the conditions on preference postulated by the theorem are construed as conditions of rationality or conditions of measurability. In the former case the theorem can be interpreted as showing that acceptance of the rationality conditions on preference commits one to the theory's claims about the properties of rational partial belief and desire. While in the latter case the theorem can be interpreted as showing that satisfaction of the conditions in question makes empirical measurement of partial belief and desire possible.

Now from the point of view of the problem of justifying the decision theory he invokes, Ramsey's representation theorem is not particularly helpful. For one is very unlikely to accept his axioms as definitive of rational preference for conditional prospects unless one accepts the theory of expected utility that motivates them. This is particularly true of axiom R4, which seems to have no justification other than that it secures the meaningfulness of value differences. Taken as axioms of measurement, however, they do much better for they specify in a precise way the conditions under which measures of the agent's degrees of belief and desire are determined by her choices amongst prospects.

With respect to problem of justification, on the other hand, Savage's (1954) theory is a good deal more impressive. Savage chooses his axioms of preference with an eye to their independent plausibility as rationality conditions. In-

[5] For instance, Fishburn (1981) rejects Ramsey-type theories in favour of Savage-type ones on the grounds of his 'restricted act space'. In fact, however, his set of conditional prospects is roughly equivalent to Savage's set of acts.

dependent, that is, of the quantitative theory of belief and desire that he will derive from them. Such a claim can justifiably be made for the Sure-Thing principle, for instance. One need not grant much plausibility to expected utility theory to grant that of two actions that yield the same outcomes when C is the case, one should choose the one with the preferred outcome when C is not. But Savage builds a very strong and implausible assumption into the very framework of his decision theory. He assumes that the desirability of any possible outcome of an action is independent of the state of the world in which it is realised.

Let us start by getting a general idea of the problem. It is a banal fact about our attitudes to many things that they depend on all sorts of contextual factors. Hot chocolate can be delightful on a cold evening, but sickly in the heat of a summer's day. Swimming on the other hand is best reserved for those hot days. I shall say, somewhat barbarically, that the swimming or drinking hot chocolate is desirabilistically dependent on the weather. Many things, on the other hand, are to all practical purposes desirabilistically independent, certainly swimming and the temperature on the moon are for me. Any reasonable theory of rational agency ought to recognise these banal facts.

How does Savage's theory violate them? Savage uses observations of choices amongst actions to determine agents' attitudes. Actions, on his account, are functions from states of the world to possible outcomes: when you choose an action you choose to make it true that if the world is in state s_1 then outcome o_1 will be realised, if it is in state s_2 then outcome o_2 will be realised, and so on. Now if we are to recognise that the desirability of the outcomes of actions depend on the state of affairs in which they are realised, then either the utilities we derive for them must be state-dependent i.e. of the form $U(o_2|s_1)$, or the outcome o_1 must include the fact that s_1 prevails (as outcomes of Ramsey's conditional prospects do). But Savage both assumes that any combination of state and outcome is possible and assigns state-independent utilities to outcomes.[6]

On our reconstruction of Ramsey, not all outcomes are achievable in any given state. So his theory requires no violation of the banal facts concerning the interdependence of our attitudes. Instead of building desirabilistic independence into his framework, he postulates the existence of only a very limited class of possibilities – those represented by ethically neutral propositions – which are desirabilistically independent of all others. One may question whether there are any propositions that are truly ethically neutral, but there are

[6] The problem is in fact a good deal more complicated than my presentation suggests, but not in ways that matter here. For a more detailed discussion see Schervish et al (1990).

clearly some that are good approximations. The postulation of their existence is not a heavy burden for such an idealised account to bare.

This is not, of course, the end of the matter. There have been numerous attempts to solve the problem of state-dependent utilities (as it has become known) within Savage's framework. Many of the proposed solutions are ingenious, but they always come at the cost of greater complexity and more burdensome assumptions. This is not the appropriate place to review the literature, but anyone who has ploughed through it will have little difficulty in recognising the merits of the elegantly simple method that Ramsey invented.

5. Jeffrey's Objection

In motivating his own method of measuring belief, Ramsey argues that the established method of offering bets with monetary rewards to someone to elicit their degrees of belief is 'fundamentally sound' but suffers from being both insufficiently general and necessarily inexact. Inexact partly because the marginal utility of money need not be constant, partly because people may be especially averse (or otherwise) to betting because of the excitement involved and partly because "the proposal of a bet may … alter his state of opinion" (Ramsey 1926, 68)

Ramsey seems to think that his own theory is not vulnerable to these problems, even though his method is similar in many ways to the one he is criticising. Not everyone would agree. Richard Jeffrey (1983), for instance, has argued that just such a problem plagues Ramsey's own account. In order to measure agents' partial beliefs, Ramsey requires that they treat possibilities like it being α if P and β if not as real prospects i.e. things that can be brought about by choice. But to persuade someone of our power to bring it about at will that it will be sunny tomorrow if the coin lands heads and snowy if it lands tails is to cause them to entertain possibilities which at present they do not. That is one must modify their beliefs in order that one may better measure them! There is of course no guarantee then that the measurements so effected are not, at least partially, artifacts of the measurement process itself.

How might Ramsey respond to this problem? A natural response to Jeffrey's objection would be to say that Ramsey does not, in fact, require that agents really believe in such fanciful causal possibilities. All that he requires is that they choose amongst gambles *as if* they believed that they would truly yield the advertised consequences under the relevant conditions. To be sure, such a response will not satisfy a behaviourist, for introspection on the part of agents must then play a crucial role in the production of their choices. For when we ask the agent to choose between an prospect which yields sunny weather if Labour wins the next election and rainy weather if they do not, and one which yields rainy weather if Labour wins the next election and sunny weather if they do not, we are in effect asking them to determine for themselves what they would prefer in the event that such gambles were reliable. But then we may as well just ask them what they would prefer and forget about the observation of choices altogether.

And indeed why not? Let us see what such a reconstrual of Ramsey's method would amount to in the context of the experimental determination of a subject's degrees of belief and desire, by comparing the following measurement schemes:

1. *Scheme A:* The subject introspects her degrees of belief and desire and then reports them to the observer.
2. *Scheme B:* The observer presents the subject with a number of options and her choice is recorded. The set of options offered is varied until a ranking over all of them has been constructed from the observations of her choices. This ranking is then used to construct a quantitative representation of her degrees of belief and desire.
3. *Scheme C:* The observer questions the subject as to which of various possibilities she would prefer were the true one. Her answers are then used to construct a ranking of all possibilities and this in turn determines a quantitative representation of her mental attitudes.

Scheme A is the method criticised by behaviourists and Ramsey alike for its naive dependence on introspection. Scheme B summarises the behaviourist's method, Scheme C the alternative interpretation of Ramsey's method. Both are underwritten by the representation theorems of Decision Theory. In Scheme C introspection plays an essential role: to provide answers to the experimenter's questions the subject must reflect upon and judge her own preferences. In Scheme B, on the other hand, though it is conceivable that the subject arrives at a choice via an introspection on her preferences, she need not do so. She may simply choose without reflection, indeed without even having the concept of preference. Scheme C is a method intimately tied to the possibility of linguistic communication and the kind of self-consciousness that typically accompanies it; Scheme B is just as applicable to earthworms as to philosophers.

I see no reason why Ramsey should be resistant to this interpretation of his method as a version of Scheme C. Although it requires him to disavow the behaviourist pretension that introspection can be completely eliminated in favour of rich observations of behaviour, it does not commit him to the view instantiated in Scheme A (and which he clearly rejects) that partial beliefs and desires can be directly introspected. In this sense this interpretation does not conflict with anything that he says. And it has the crucial advantage of extricating him from Jeffrey's objection.

REFERENCES

BRADLEY, R. W. 2001, "Ramsey and the Measurement of Belief" in D. Corfield and J. Williamson, eds., *Foundations of Bayesianism*, Dordrecht: Kluwer, pp. 263-290

BOLKER, E. D. 1967 "A Simultaneous Axiomatisation of Utility and Subjective Probability" *Philosophy of Science* 34, pp. 292-312

DAVIDSON, D. and SUPPES, P. 1956 "A Finitistic Axiomatisation of Subjective Probability and Utility" *Econometrica* 24, pp. 264-275

DE FINETTI, B. 1937 "La Prévision: ses lois logiques, ses sources subjectives" *Annales de l'Institut Henri Poincaré*, vol. 7

FISHBURN, P. C. 1981 "Subjective Expected Utility: A Review of Normative Theories" *Theory and Decision* 13, pp. 139-199

JEFFREY, R.C. 1983 *The Logic of Decision*, 2nd edn, Chicago: University of Chicago Press

KRANTZ, D. H., R. DUNCAN LUCE, P. SUPPES and A. TVERSKY 1971 *Foundations of Measurement*, vol.1, London: Academic Press

PFANZAGL, J. 1968 *Theory of Measurement*, New York: Wiley

RAMSEY, F. P. 1926 "Truth and Probability" in D. H. Mellor, ed., *Philosophical Papers*, Cambridge: Cambridge University Press, 1990.

SAHLIN, N.-E. 1990 *The Philosophy of F. P. Ramsey*, Cambridge: Cambridge University Press.

SAVAGE, L. J. 1954 *The Foundations of Statistics*, 2nd edition, New York: Dover, 1972

SCHERVISH, M. J., SEIDENFELD T. and KADANE J.B. 1990 "State-dependent Utilities" *Journal of the American Mathematical Association* 85, pp. 840-847

SOBEL, J. H. 1998 "Ramsey's Foundations Extended to Desirabilities" *Theory and Decision* 44, pp. 231-278