Ramsey and the Ethically Neutral Proposition

Edward Elliott

Australian National University

In his posthumously published 'Truth and Probability', Frank Ramsey sketches a proposal for the empirical measurement of credences, along with a corresponding set of conditions for a (somewhat incomplete) representation theorem intended to characterize the preference conditions under which this measurement process is applicable. Ramsey's formal approach is distinctive, deriving first a utility function to represent an agent's utilities, and then using this to construct their credence function. In specifying his measurement process and his conditions, however, Ramsey introduces the notion of an *ethically neutral proposition*, the assumed existence of which plays a key role throughout Ramsey's system.

The existence of such propositions has often been called into question. Ramsey's own definition of ethical neutrality presupposes the philosophically suspect theory of logical atomism. On other common ways of defining the notion, it's frequently noted that we lack good reasons for supposing that ethically neutral propositions exist, and in some cases we find that there are very good reasons for supposing that they cannot exist. Any system which relies on the existence of such propositions ought to be rejected.

In this chapter, I will first outline Ramsey's proposal in some detail (§1), highlighting some empirical and exegetical difficulties along the way. This in-depth look at the proposal will help us to see why Ramsey thought he needed to introduce the notion of ethical neutrality (§2.1), and why any theorem which appeals to ethically neutral propositions should be considered highly problematic (§2.2).[1]

1. Ramsey's proposal

One of Ramsey's main goals in 'Truth and Probability' was to argue that the laws of probability supply for us a "logic of partial belief'" (1931, 166), much as the laws of deductive logic might be taken to supply for us a logic of full belief. The overall argument of the paper proceeds by first supplying an account of what credences *are*, and on the basis of that account, showing that credences *must* be probabilistically coherent.[2]

Regarding the first step, of defining credences, Ramsey clearly had operationalist sympathies, asserting that the notion "has no precise meaning unless we specify more exactly how it

[1] This is a lightly modified version of Chapter 7 of my PhD thesis, *Representation Theorems and the Grounds of Intentionality*. It also largely overlaps with the first half of my paper, 'Ramsey without Ethical Neutrality', forthcoming in *Mind*. In both works, I show how to construct a small number of related representation theorems which are essentially Ramseyean in character (i.e., with similar formal primitives and proof structure), but which do not require positing ethically neutral propositions in any problematic sense.

[2] Ramsey frequently changes between descriptive language and normative language and seems to treat his claims about our beliefs and preferences as both descriptively accurate and normatively compelling. As a result, it is very unclear whether he took probabilistic coherence to be something that actually holds true of ordinary agents, or as an ideal that ordinary agents ought to aspire to. Each reading of the text presents its own difficulties, so I have opted to remain neutral on the matter.

is to be measured" (1931, 167). To be measured as having probabilistically coherent credences *is* (more or less), on this picture, to have probabilistically coherent credences, and anyone who can be measured through Ramsey's procedure at all will have credences conforming to the laws of probability.

Setting operationalism aside, it's easy to see in 'Truth and Probability' an early statement of a functionalism based on preferences: credences are to be understood through their role with respect to preferences when considered in conjunction with a total utility state. Ramsey writes that "the degree of a belief is a causal property of it, which we can express vaguely as the extent to which we are prepared to act on it" (1931, 169). Ramsey argues against characterising credences in terms of some introspectively accessible feeling of confidence had by a subject upon considering the relevant proposition. These arguments go well beyond any brute or simplistic version of operationalism, though I will not recapitulate them here. He concludes that "intensities of belief-feelings … are no doubt interesting, but … their practical interest is entirely due to their position as the hypothetical causes of beliefs *qua* bases of action" (1931, 172). On this more charitable interpretation, Ramsey's representation theorem can be seen as spelling out precisely the relevant functional roles associated with credences at least partly in terms of their connection with preferences.

Ramsey proposes to take as the theoretical basis of his measurement system a particular theory of decision making—that is, the theory that "we act in the way that we think most likely to realize the objects of our desires, so that a person's actions are completely determined by his desires and opinions" (1931, 173). His idea was to assume the basic truth of something like standard subjective expected utility theory, and *given that assumption*, use empirical information about an agent's preferences to work out what her credences and utilities must be. Ramsey was entirely aware of the empirical difficulties facing the theory, writing that:

> [it] is now universally discarded, but nevertheless comes, I think, fairly close to the truth in the sort of cases with which we are most concerned … This theory cannot be made adequate to all the facts, but it seems to me a useful approximation to the truth particularly in the case of our self-conscious or professional life, and it is presupposed in a great deal of our thought. (1931, 173)

We will return shortly to what Ramsey meant by "the sort of cases with which we are most concerned", and exactly what he needed to assume to get his measurement process off the ground.

At several points Ramsey notes that similar 'approximations to the truth' are used frequently by physicists and other hard scientists in the development of systems for the measurement of non-psychological quantities. Measurement systems are never developed in a theoretical vacuum; the actual data we receive is always interpreted through the lens of one presupposed theory or another—and quite frequently, as Ramsey notes, such a theory might 'like Newtonian mechanics … still be profitably used even though it is known to be false' (1931, 173), so long as it is accurate *for the cases at hand*. Indeed, at several points Ramsey argues that measurement 'cannot be accomplished without introducing a certain amount of hypothesis or fiction … [and] if it is allowable in physics it is allowable in psychology also' (1931, 168; cf. Krantz *et al* 1971, 26–31 on the role of idealizing assumptions in the construction of measurement systems). For Ramsey, this hypothesis or fiction is that ordinary folk are more or less rational expected utility maximisers, at least under in the right circumstances and with respect to the right kinds of decision situations.

With this in mind, we can summarise Ramsey's measurement procedure as follows:

(1) Determine *S*'s preferences over outcomes and gambles
(2) Define a relation of equal difference in utilities
(3) Locate ethically neutral propositions of credence ½
(4) Construct an interval scale representation $\mathcal{Des}$ of *S*'s preferences
(5) Use $\mathcal{Des}$ to define a probability function $\mathcal{Bel}$

I will discuss each step in turn. For the sake of simplicity, I have neglected to discuss one important aspect of Ramsey's procedure: the use of preferences over complex gambles to define conditional probabilities, which are necessary to ensure that the measured credences constitute a probability function. This part of Ramsey's procedure is outlined in (Bradley 2001). It is also worth noting that much of what follows is a rational reconstruction—Ramsey's own remarks are sketchy at best, and he rarely explains the motivations for any of the moves he makes.

1.1 Determining a preference ordering over outcomes and gambles

The very first stage of Ramsey's procedure is to determine the agent's basic preferences over different ways the world might be. This is, he says, relatively straightforward:

> If … we had the power of the Almighty, and could persuade our subject of our power, we could, by offering him options, discover how he placed in order of merit all possible courses of the world. In this way all *possible worlds* would be put in an order of value … (1931, 176, emphasis added)

Ramsey writes that he intends the relevant objects of preference to be "different possible totalities of events … the ultimate organic unities" (1931, 177-8)—i.e., possible worlds. I will use $\mathcal{O} = \{o_1, o_2, \ldots\}$ to designate the set of these "possible totalities of events", which for reasons of generality I'll refer to as *outcomes*.

Importantly, within only a few paragraphs, Ramsey seems to contradict his characterisation of outcomes as possible worlds by asserting that with respect to *at least one* proposition *P* and *some* pair of outcomes o_1, o_2, '[o_1] and [o_2] must be supposed so far undefined as to be compatible with both *P* and ¬*P*' (1931, 178). The most natural interpretation of this seems to be that in some select few circumstances at least, o_1 and o_2 ought to be considered *not quite* as worlds, but rather as propositions maximally specific with respect to everything *except P*. The reason for this apparent change of heart will be discussed in some detail below; for now, we will treat $\mathcal{O}$ as a set of very highly specific consistent propositions, some—but *not all*—of which may be maximally specific.[3]

[3] It is unclear how propositions as highly specific as Ramsey suggests is outcomes to be can be *offered* to any ordinary human subject; the power to conceptualise even one possible world in all its detail is well beyond the average person. However, Ramsey nowhere explains or justifies his reasons for using possible worlds (or near-worlds) as outcomes, and with some significant tweaking this part of his system can be removed so that outcomes need not be very fine-grained at all. See (Elliott MS-a) for details.

Given a preference ordering over $\mathcal{O}$, we are also required to empirically determine how our subject ranks *gambles*.[4] Once again, Ramsey asks us to imagine that we have "persuaded our subject of our power", but this time we make offers of the following kind: "Would you rather have world [o_3] in any event, or world [o_1] if P is true, and world [o_2] if P is false?" (1931, 177). Let us represent the latter option, the gamble o_1 *if P is true,* o_2 *otherwise*, as simply $(o_1, P; o_2)$. Ramsey then notes that:

> If … [the agent] were certain that P was true, he would simply compare [o_1] and [o_3] and choose between them as if no conditions were attached; but if he were doubtful his choice would not be decided so simply. (1931, 177)

Here, Ramsey looks to be comparing an outcome with a gamble, so we are to assume that gambles and outcomes are comparable. It is also evident from the discussion that follows that we need to consider agents' preferences between gambles. In sum: if $\mathcal{G}$ is a set of *two-outcome* gambles of the form $(o_1, P; o_2)$ where o_1 and o_2 are both members of $\mathcal{O}$, and $x \succcurlyeq y$ represents a "weak preference" for x over y—i.e., either x is strictly preferred to y ($x \succ y$) or they are equi-preferable ($x \sim y$)—then Ramsey requires us to empirically determine a complete weak preference ordering $\succcurlyeq$ on $\mathcal{O} \cup \mathcal{G}$.[5]

There are a number of difficulties with Ramsey's proposal that might be raised at this point; I will discuss two. The first is exegetical, and concerns the notion of a *gamble* and how we are to understand *preferences* over them. On a natural reading, a gamble like "world [o_1] if P is true, and world [o_2] if P is false" is an offer made by another agent, one which a decision-maker may choose to accept or reject. Accepting the offer is then a very specific kind of act, which—depending on the capacities and trustworthiness of the offer-maker—will result in such-and-such an outcome should some proposition turn out to be true or false.[6]

It would be a mistake, I think, to interpret the relation $(o_1, P; o_2) \succ (o_3, Q; o_4)$ in the decision-theorists' standard sense as a disposition to accept a gamble which would *in fact* result in o_1 *if P is true,* o_2 *otherwise*, where the only other offer available a gamble which would *in fact* result in o_3 *if Q is true,* o_4 *otherwise* (and where exactly one of the offers must be accepted). This way of interpreting $\succcurlyeq$ on $\mathcal{O} \cup \mathcal{G}$ does not take into account the way in which the decision-maker *conceives* of the two options available to her. Dispositions to choose between pairs of offers will depend on how the agent in question conceives of the options available, and there is no guarantee that by simply offering S a gamble that *as a matter of fact* has the outcomes o_1 if P is true, o_2 otherwise, that S will therefore *represent* that gamble as such to herself—S may

[4] In many discussions, what I will call here 'gambles' are also sometimes called 'bets'; the difference in terminology is purely a matter of choice and not intended to signify anything important.

[5] A preference ranking $\succcurlyeq$ on some set $\mathcal{X}$ is *complete* just in case, for all members $x, y \in \mathcal{X}$, either $x \succcurlyeq y$ or $y \succcurlyeq x$ (or both).

[6] Contrary to what is frequently claimed, Ramsey does not posit his preference relation to hold between (formal representations of) *arbitrary* acts that an agent might make in their present situation—indeed, his simple two-outcome gambles lack sufficient structure to usefully represent the vast majority of ordinary acts, as the discussion that follows should make clear. At best, they might be taken to represent very special kinds of acts, made in very specific circumstances.

have misheard, or may not trust the offer-maker's willingness or capacity to make good on the offer.

Suppose, by way of example, that S is offered a gamble Γ, where the offer-maker claims that by accepting Γ she will win \$100 if P is true and lose \$10 otherwise. As a matter of fact, the offer-maker is telling the truth: that really is what would happen should S accept the offer. However, S also has evidence that the offer-maker sometimes get the prizes mixed up, and so she have some credence that accepting Γ she would lose \$10 just in case P is true and win \$100 just in case $\neg P$ is true. In this situation, it would be a mistake to represent the act of *accepting* Γ as (+\$100 if P; -\$10 if $\neg P$), although this does in fact capture its modal profile under one very coarse-grained description. The problem, of course, is that the representation—while accurate to the facts of the matter—fails to capture how the decision-maker conceptualises the act in terms of her uncertainty regarding what outcomes might obtain should she perform it. Instead, a better representation would be the more fine-grained (+\$100 if *P and the offer is not mixed up*; -\$10 if *¬P and the offer is not mixed up*; -\$10 if *P and the offer is mixed up*; +\$100 if *¬P and the offer is mixed up*). In this case, the choice she makes will not depend solely on her credences with respect to P and $\neg P$, but also her credences with respect to whether the offer is mixed up or not.

Some recent commentators have chosen to forego the choice-theoretic interpretation of Ramsey's preference relation and instead treated preferences between his gambles as a very particular subset of our more ordinary preferences *qua* mental states, where these are to be understood not in terms of choice dispositions but as a kind of relational propositional attitude. On this understanding, for any two propositions P and Q, $P \succcurlyeq Q$ just in case one attaches at least as much value to P being true as to Q being true. In this vein, Sobel (1998, 239) suggests that Ramsey's gamble $(o_1, P; o_2)$ can be treated as just a conjunction of counterfactual conditionals,

$$(P \mathbin{\Box\!\!\longrightarrow} o_1) \,\&\, (\neg P \mathbin{\Box\!\!\longrightarrow} o_2)$$

The idea, apparently, is that by accepting the offer of a gamble that returns o_1 *if P is true*, o_2 *otherwise*, one makes that pair of counterfactual conditionals *true* (at least under the assumption that the offer-maker is capable and willing to make good on the offer). Thus, one will choose $(o_1, P; o_2)$ over $(o_3, Q; o_4)$ just in case one would prefer the truth of $(P \mathbin{\Box\!\!\longrightarrow} o_1) \,\&\, (\neg P \mathbin{\Box\!\!\longrightarrow} o_2)$ to the truth of $(Q \mathbin{\Box\!\!\longrightarrow} o_3) \,\&\, (\neg Q \mathbin{\Box\!\!\longrightarrow} o_4)$. Somewhat more plausibly, a disposition to choose one offered gamble over another will sometimes *go hand in hand with* a preference for the truth of some conjunction of counterfactuals over another—namely, that conjunction which the agent believes would *very likely* be made true by her choice (and not necessarily the pair of conditionals which *would* be made true by her choice). In either case, the upshot on this picture is that we can do away with talk of dispositions to choose between offered gambles and simply construe Ramsey's preferences as standing propositional attitudes towards pairs of propositions of a very specific form. In most cases, these attitudes will correlate with a subject's choice dispositions, but they hold independently of and presumably prior to those dispositions.

On the other hand, Bradley (1998, 193-4) rejects the use of counterfactuals and instead treats his Ramsey-style gambles as conjunctions of *indicative* conditionals. Because he also wants to accept Adams' Thesis (see Adams 1975), he explicitly foregoes any propositional interpretation of $(o_1, P; o_2)$, and instead treats the members of his set $\mathcal{G}$ as sentences in a formal language which do not correspond to any particular collection of possible worlds. It is unclear, however, what these preferences over *sentences* could amount to—I attach value to different ways the world might be, not to uninterpreted sentences.[7]

I will not say much more about how Ramsey's gambles are to be interpreted, except to note that on any of the foregoing suggestions there will be problems with Ramsey's proposal *qua* procedure for the empirical measurement of credences and utilities. Without 'peaking inside the head', we can from the outside characterise the simple act of *accepting a gamble* in terms of its *actual* modal profile, and understand preferences between them in terms of choice dispositions. But such a characterisation is liable to lead to mis-measurements whenever the subject is uncertain or misled about the outcomes associated with any gamble she accepts. (In the example just above, S could accept Γ because she wants \$100 and things P is highly likely, *or* because she wants \$100 and thinks $\neg P$ is really likely but also that the offer-maker mischaracterised the gamble's payoffs.) On the other hand, we might skip the more usual choice-theoretic characterisation of preferences and their objects and consider the subject's mental attitudes directly—but, in doing so, we would be helping ourselves to quite a lot of the information that the procedure itself was supposed to supply us with.

The second difficulty for the proposed procedure is that convincing a subject that "we had the power of the Almighty" would surely drastically alter her doxastic state prior to measuring it and thus unduly influence the state to be measured, as Jeffrey (1983, 158-60) has noted. For instance, ordinary people don't usually believe that those they meet are omnipotent, and making someone come to believe otherwise would likely causes changes to a lot of their worldview! A small change in the quantities being measured is to be expected with any measurement procedure—a thermometer tends to either cool or heat its surroundings by a very small amount whenever it is used, and this is hardly problematic—but changes as drastic as this create more serious problems of accuracy. A related problem is that when a subject is offered the choice between (say) either o_3 or $(o_1, P; o_2)$, we must not suppose that her credence in P is in any way changed by the offer, or this would ruin the measurement. Nevertheless, we could easily imagine some propositions for which this kind of change would surely happen—including *that psychologists are capable of making arbitrary propositions come true*.

Interestingly, Ramsey himself objects to the standard *betting interpretation* of credences on the grounds that "the proposal of the bet may inevitably alter [the subject's] state of opinion" (1931, 172). Either Ramsey did not recognise that the same objection applies with greater force to his own account, or he believed that the worry could be addressed. Bradley (2001, 285-8) suggests one way in which the worry might be addressed: rather than making the subject believe in our godlike powers, we simply ask her to judge her preferences amongst options *as if* they were genuinely available to her. To the extent that such a request can be satisfied, this re-construal of Ramsey's methodology may help to minimise any changes to subjects' credences prior to measurement.

[7] For reasons outlined by Joyce (1999, 62-3) and omitted here, "o_1 if P is true, and o_2 if P is false" cannot plausibly be understood using material conditionals without destroying the distinction between importantly different gambles.

1.2 Defining an equal difference relation

Ramsey's first step has us empirically determine how the agent ranks outcomes and gambles. However, a simple preference ordering on outcomes and gambles only suffices for an ordinal scale representation of an agent's utilities for those outcomes and gambles. For Ramsey, this is unsatisfactory: "There would be no meaning in the assertion that the difference in value between [o_1] and [o_2] was equal to that between [o_3] and [o_4]" (1931, 176). Thus Ramsey sets himself the task of characterizing an *equal difference* (in utilities) relation between pairs of outcomes wholly in terms of preferences over gambles. If he can do this, then on the basis of well-known results from the mathematical theory of measurement, he can construct a richer representation of our utilities—one which is capable of representing the different *strengths* with which we desire the different outcomes.

Let us say that $(o_1, o_2) =^{d} (o_3, o_4)$ holds iff the difference in value for the agent between o_1 and o_2 is equal to the difference in value between o_3 and o_4. Ramsey's goal of defining $=^{d}$ in terms of preferences over gambles then sets up a certain difficulty to be overcome. According to the background assumption that agents are good expected utility maximisers, an agent's preferences over gambles are determined by two factors: their *utilities* and their *credences*. Whether an agent prefers $(o_1, P; o_2)$ to $(o_3, Q; o_4)$, for example, depends partly on the utilities that she attaches to o_1, o_2, o_3, o_4, and partly on the credences regarding P and Q. However, whether $(o_1, o_2) =^{d} (o_3, o_4)$ holds for that agent should depend *solely* on the utilities she attaches to o_1, o_2, o_3, o_4. In order to define $=^{d}$ in terms of preferences over gambles, then, Ramsey needs some way of factoring out any confounding influences, so that whether the agent prefers one of the relevant gambles to another depends *only* on the utilities attached to the outcomes involved.

Ramsey's central innovation here is to define, in terms of preference, what it is for an agent to have credence ½ in a proposition, and then to use this to define $=^{d}$. Let us suppose for now that whether an agent prefers $(o_1, P; o_2)$ to $(o_3, Q; o_4)$ depends *only* on the utilities the agent has for o_1, o_2, o_3, o_4, and the credences she has for P and Q. More specifically, assume *Naïve Expected Utility Theory*:

Naïve Expected Utility Theory

If $\mathcal{D}es$ is a real-valued function that models the agent's utilities, and $\mathcal{B}el$ is a credence function that models the agent's credences, then $(o_1, P; o_2) \succcurlyeq (o_3, Q; o_4)$ iff $\mathcal{D}es(o_1).\mathcal{B}el(P) + \mathcal{D}es(o_2).(1 - \mathcal{B}el(P)) \geq \mathcal{D}es(o_3).\mathcal{B}el(Q) + \mathcal{D}es(o_4).(1 - \mathcal{B}el(Q))$

We will note shortly that Ramsey did *not* assume Naïve Expected Utility Theory; but for now it suffices to explain the reasoning behind his definitions. It is worth noting that $\mathcal{B}el$ is not here assumed to be a probability function as opposed to any other kind of function that takes us from propositions into the [0, 1] interval, though it is implicit in Naïve Expected Utility Theory that $\mathcal{B}el(\neg P) = 1 - \mathcal{B}el(P)$.[8] Were this *not* the case, we would not expect the contribution of o_2 to the desirability of $(o_1, P; o_2)$ to be determined by $\mathcal{D}es(o_2).(1 - \mathcal{B}el(P))$ as opposed to $\mathcal{D}es(o_2).\mathcal{B}el(\neg P)$ directly.

[8] A probability function $\mathcal{P}r$ is a mapping from a set $\boldsymbol{\mathcal{P}}$ of propositions into the [0, 1] interval which obeys the following constraints: (i) $\boldsymbol{\mathcal{P}}$ should be closed under negation and disjunction, (ii) where $\top$ is necessary, $\mathcal{P}r(\top) = 1$ and $\mathcal{P}r(\neg\top) = 0$, and (iii) $\mathcal{P}r$ must be *additive*, in the sense that for any two mutually exclusive $P, Q \in \boldsymbol{\mathcal{P}}$, $\mathcal{P}r(P \vee Q) = \mathcal{P}r(P) + \mathcal{P}r(Q)$.

Suppose that the agent is indifferent between $(o_1, P; o_2)$ and $(o_2, P; o_1)$. According to Naïve Expected Utility Theory, there are only two (not mutually exclusive) ways in which this might come about: either both o_1 and o_2 have exactly the same utility for the agent, or the agent's credence in P is exactly ½. To rule out the former possibility, we consider a pair of gambles $(o_3, P; o_4)$ and $(o_4, P; o_3)$, where we know that the agent is not indifferent between o_3 and o_4. This we will have already determined by considering the agent's preferences over outcomes. If we then find that the agent is indifferent between $(o_3, P; o_4)$ and $(o_4, P; o_3)$, we will have established that $\mathcal{B}el(P) = ½$. If her credence in P were any other way, then the agent would have not been indifferent between $(o_3, P; o_4)$ and $(o_4, P; o_3)$.

With this in place, Ramsey notes that we are then able to say that $(o_1, o_2) =^{d} (o_3, o_4)$ holds iff $(o_1, P; o_4) \sim (o_2, P; o_3)$, where P is such that the agent believes it to degree ½. The reasoning behind this is not immediately obvious, so it will be worth spelling out. From the assumption of Naïve Expected Utility Theory, we have that $(o_1, P; o_4) \sim (o_2, P; o_3)$ holds just in case:

$$\mathcal{D}es(o_1).\mathcal{B}el(P) + \mathcal{D}es(o_4).(1 - \mathcal{B}el(P)) = \mathcal{D}es(o_2).\mathcal{B}el(P) + \mathcal{D}es(o_3).(1 - \mathcal{B}el(P))$$

We have also already established that $\mathcal{B}el(P) = ½ = 1 - \mathcal{B}el(P)$, so we can drop the constant factor leaving us with:

$$\mathcal{D}es(o_1) + \mathcal{D}es(o_4) = \mathcal{D}es(o_2) + \mathcal{D}es(o_3)$$

Which holds just in case:

$$\mathcal{D}es(o_1) - \mathcal{D}es(o_2) = \mathcal{D}es(o_3) - \mathcal{D}es(o_4)$$

This just states that the difference between o_1 and o_2 is equal to the difference between o_3 and o_4; so if $\mathcal{B}el(P) = ½$, $(o_1, P; o_4) \sim (o_2, P; o_3)$ iff $(o_1, o_2) =^{d} (o_3, o_4)$.

1.3 Locating ethically neutral propositions

Before moving on to measuring utilities, however, Ramsey makes the following note:

> There is first a difficulty which must be dealt with; the propositions like P … which are used as conditions in the [gambles] offered may be such that their truth or falsity is an object of desire to the subject. This will be found to complicate the problem, and we have to assume that there are propositions for which this is not the case, which we shall call ethically neutral. (1931, 177)

This is the entirety of what Ramsey writes regarding his motivation for introducing ethically neutral propositions.

The idea is clear enough: Naïve Expected Utility Theory is mistaken, as it fails to take into account the utility that may attach to the gamble's condition and how the condition might influence the agent's valuation of the elements of $\mathcal{O}$. Assuming that o_1 is consistent with both P and $\neg P$, it's possible that an agent might attach a different value to $(o_1 \& P)$ than to $(o_1 \& \neg P)$. These are potentially quite different states of affairs with potentially different utilities, and the truth or falsity of P might make a great deal of difference to how the outcome o_1 is valued. For instance, suppose that in o_1 the agent has a puppy as a pet, while in o_2 she instead keeps a kitten, and let P be *puppies spread disease but kittens don't*; plausibly, $(o_1 \& P)$ will be valued quite differently than $(o_1 \& \neg P)$, and likewise for $(o_2 \& P)$ and $(o_2 \& \neg P)$.

Instead of Naïve Expected Utility Theory, and supposing o_1, o_2, o_3, and o_4 are each compatible with the relevant propositions, we should really have that:

$$(o_1, P; o_2) \succcurlyeq (o_3, Q; o_4)$$

Just in case:

$$\mathcal{D}es(o_1 \& P).\mathcal{B}el(P) + \mathcal{D}es(o_2 \& \neg P).(1 - \mathcal{B}el(P))$$

Is at least as great as:

$$\mathcal{D}es(o_3 \& Q).\mathcal{B}el(P) + \mathcal{D}es(o_4 \& \neg Q).(1 - \mathcal{B}el(Q))$$

It is easy to see that this fact invalidates the reasoning behind both the definition of what it is for an agent to have a credence ½ in a proposition, and the definition of $=^d$, for now we can no longer say that the agent's preferences between $(o_1, P; o_2)$ and $(o_3, Q; o_4)$ depend on their credences in P and Q and the utilities the agent has for o_1, o_2, o_3, o_4. Rather, they actually depend on the agent's credences in P and Q and utilities for $(o_1 \& P)$, $(o_2 \& \neg P)$, $(o_3 \& Q)$, and $(o_4 \& \neg Q)$.

Ramsey's solution to this difficulty is the ethically neutral proposition—a kind of proposition the truth or falsity of which is of absolutely no concern to the agent. Ramsey provides us with a problematic definition of the notion, which I will discuss further in §2.2. The evident purpose of its introduction, however, is that if P is ethically neutral, then the conjunction of P with o has the same utility as o itself, and similarly for the conjunction of $\neg P$ and o. Setting aside Ramsey's own definition, then, we can say that P is *ethically neutral* whenever $o \sim (o \& P) \sim (o \& \neg P)$, for any $o \in \mathcal{O}$ that is compatible with both P and $\neg P$.

So long as we are considering gambles conditional on ethically neutral propositions, we can *without risk of making the above kind of error* apply Naïve Expected Utility Theory. Thus Ramsey happens upon the following two definitions:

Definition 1: Ethically neutral proposition of credence ½
P is an ethically neutral proposition of credence ½ iff P is ethically neutral, and for some o, $o_2 \in \mathcal{O}$, $\neg(o_1 \sim o_2)$, and $(o_1, P; o_2) \sim (o_2, P; o_1)$

And:

Definition 2: Equal difference relation
$(o_1, o_2) =^d (o_3, o_4)$ iff $(o_1, P; o_4) \sim (o_2, P; o_3)$, where P is an ethically neutral proposition of credence ½

1.4 Measuring utilities

At this point, Ramsey lists eight preference conditions, and states (but does not prove) that their satisfaction enables an appropriately rich representation of the agent's preferences. Let $\mathcal{P}$ be a set of propositions, $\mathcal{O}$ the set of outcomes, and $\mathcal{G}$ the set of gambles; $\succcurlyeq$ is defined on $\mathcal{O} \cup \mathcal{G}$. Ramsey's Representation Conjecture can then be stated thus:

Ramsey's Representation Conjecture

If **RAM1–8** hold of $<\mathcal{P}, \mathcal{O}, \mathcal{G}, \succcurlyeq>$, then there exists a real-valued function $\mathcal{D}es$ on $\mathcal{O}$ such that for all $o_1, o_2, o_3, o_4 \in \mathcal{O}$,

$$\mathcal{D}es(o_1) - \mathcal{D}es(o_2) = \mathcal{D}es(o_3) - \mathcal{D}es(o_4) \text{ iff } (o_1, o_2) =^d (o_3, o_4)$$

Furthermore, $\mathcal{D}es$ is unique up to positive linear transformation

We will not consider whether Ramsey's preference conditions successfully ensure the desired representation result, or how they might be fleshed out to do so if not—though see (Bradley 2001) and (Elliott forthcoming) for relevant work in this regard. It is clear that something in the vicinity of Ramsey's conditions should suffice, though I will not take a stand on the precise formulation needed.[9]

The very first preference condition is the most distinctive aspect of Ramsey's theorem:

RAM1 There is at least one ethically neutral proposition of credence ½

The importance of **RAM1** for the rest of Ramsey's formal system should not be understated. Most of the preference conditions to follow are stated in terms of $=^d$, which is defined in terms of ethically neutral propositions. If **RAM1** is false, those conditions will be in some cases false, in others trivial; in either case, the system as a whole collapses without this foundational assumption.

The next three preference conditions are each obviously necessary for Ramsey's desired representation result. For all $P, Q \in \mathcal{P}$, $o_1, o_2, o_3, o_4, o_5, o_6 \in \mathcal{O}$, $(o_1, P; o_2), (o_3, P; o_4) \in \mathcal{G}$, and $x, y, z \in \mathcal{O} \cup \mathcal{G}$,

RAM2 (i) If P, Q, are both ethically neutral propositions of credence ½, and $(o_1, P; o_2) \sim (o_3, P; o_4)$, then $(o_1, Q; o_2) \sim (o_3, Q; o_4)$, and (ii) if $(o_1, o_2) =^d (o_3, o_4)$, then $o_1 \succ o_2$ iff $o_3 \succ o_4$, and $o_1 \sim o_2$ iff $o_3 \sim o_4$

[9] Part of the problem with fleshing out Ramsey's conjecture is making his sets $\mathcal{O}$, $\mathcal{P}$, and $\mathcal{G}$ precise. As Ramsey outlines his system, each of these are far too sparsely characterised for the purposes of any interesting representation theorem. What we need are clear statements regarding what kinds of things can go into $\mathcal{O}$ and $\mathcal{P}$; what, if any, structural conditions they satisfy; and how exactly $\mathcal{G}$ is to be formally constructed out of $\mathcal{O}$ and $\mathcal{P}$. There are many options here, and the choice of one instead of the other can have a significant effect on how we interpret any representation result that might be proven given Ramsey's axioms (or something in the nearby vicinity). For instance, if $\mathcal{G}$ is taken as a subset of $\mathcal{O} \times \mathcal{P} \times \mathcal{O}$, then one important question concerns whether there are any restrictions on what subset that might be—such as whether it might include gambles in which the outcomes are incompatible with the conditions under which they are supposed to be won (see §2.1 for discussion). Furthermore, under this way of characterising $\mathcal{G}$, $(o_1, P; o_2)$ and $(o_2, \neg P; o_1)$ will represent distinct entities and an axiom will need to be added to Ramsey's system to ensure that they are always ranked the same. (This is necessary to establish that $\mathcal{B}el(P) = 1 - \mathcal{B}el(\neg P)$.) On the other hand, if $\mathcal{G}$ is treated instead as a collection of functions from complementary pairs of propositions in $\mathcal{P}$ into $\mathcal{O}$, then $(o_1, P; o_2)$ and $(o_2, \neg P; o_1)$ will be mere notational variants of one another, and no such additional axiom will be required.

RAM3 $\sim$ is transitive

RAM4 $=^{\mathrm{d}}$ is transitive

The role of **RAM2** is ensure that the definition of $=^{\mathrm{d}}$ is coherent—that our preferences work in such a way as required for Definition 2 to be useful. Together, **RAM2–RAM4** help to ensure that $=^{\mathrm{d}}$, which holds between pairs of outcomes, mirrors the behaviour of the *equals* relation between the differences of pairs of real numbers.

The following two existential conditions are stated in terms of what Ramsey calls *values*. Formally,

Definition 3: The value of *o*
For every $o \in \mathcal{O}$, let $\underline{o} = \{o' \in \mathcal{O}: o' \sim o\}$

The value of an outcome o, denoted $\underline{o}$, is the set of all outcomes in $\mathcal{O}$ with the same desirability as o. Ramsey's next two conditions are then:

RAM5 For all $\underline{o}_1$, $\underline{o}_2$, $\underline{o}_3$, there is exactly one $\underline{o}_4$ such that $(o_1, o_4) =^{\mathrm{d}} (o_2, o_3)$

RAM6 For all $\underline{o}_1$, $\underline{o}_2$, there is exactly one $\underline{o}_3$ such that $(o_1, o_3) =^{\mathrm{d}} (o_3, o_2)$

RAM5 implies that there is always at least one outcome o_4 such that the difference between o_1 and o_4 is equal to the difference between o_2 and o_3, for *any* choice of outcomes o_1, o_2 and o_3. In a manner of speaking, **RAM6** says that for any pair of outcomes o_1 and o_2, there is at least one outcome o_3 with a utility exactly half-way between that of o_1 and o_2. Given **RAM1** (which implies the non-triviality of $\succ$ on $\mathcal{O}$), this entails a *denseness* to the agent's preference structure—and correspondingly, that $\mathcal{O}$ is infinite.

Finally, Ramsey lists two other conditions, which are not spelled out in any detail:

RAM7 "Axiom of continuity:—Any progression has a limit (ordinal)" (Ramsey 1931, 179)

RAM8 Archimedean condition

What Ramsey intended for **RAM7** is something of a mystery. One guess (cf. Sobel 1998, Bradley 2001) would be that for every gamble $(o_1, P; o_2)$, there is an outcome o_3 such that $o_3 \sim (o_1, P; o_2)$. As I will note shortly, a condition to this effect seems to be required to ensure that every real number can be mapped to at least one outcome's value, and is important for the later derivation of credences. Interestingly, if this is the correct reading of **RAM7**, then it implies **RAM6** and renders the latter redundant.

Ramsey does not specify the character of **RAM8**, however it's easy to guess its role and thus what it ought to look like—as with other so-called Archimedean conditions in various representation theorems, it is supposed to rule out any one outcome or gamble being incomparably better or worse than another. More specifically, **RAM8** ensures that the numerical representation satisfies the Archimedean property of real numbers: for any positive number x, and any number y, there is an integer n such that $n + x \geq y$.

1.5 Measuring credences

Suppose that our subject's preferences satisfy some tidied up version of the foregoing axioms and we have our function $\mathcal{D}es$, which is unique up to positive linear transformations. Ramsey then argues that:

> Having thus defined a way of measuring value we can now derive a way of measuring belief in general. If the option of $[o_2]$ for certain is indifferent with that of $[(o_1, P; o_3)]$, we can define the subject's degree of belief in P as the ratio of the difference between $[o_2]$ and $[o_3]$ to that between $[o_1]$ and $[o_3]$ … This amounts roughly to defining the degree of belief in P by the odds at which the subject would bet on P, the bet being conducted in terms of differences of value as defined. (1931, 179-80)

In a footnote, Ramsey adds that '$[o_1]$ must include the truth of P, $[o_3]$ its falsity; P need no longer be ethically neutral' (1931, 179). We are led to the following definition:

Definition 4: Ramsey's $\mathcal{B}el$
For all contingent propositions P and outcomes o_1, o_2, o_3 such that o_1 implies P, o_3 implies $\neg P$, $\neg(o_1 \sim o_3)$, and $o_2 \sim (o_1, P; o_3)$, $\mathcal{B}el(P) = (\mathcal{D}es(o_2) - \mathcal{D}es(o_3))/(\mathcal{D}es(o_1) - \mathcal{D}es(o_3))$

Ramsey mistakenly states that Definition 4 "only applies to partial belief and does not include certain beliefs" (1931, 180), though perhaps he meant that the definition does not apply if P is *non-contingent*. In this case, we simply stipulate that $\mathcal{B}el(P) = 1$ if P is necessary, 0 if P is impossible.

The reasoning behind this final step is again left up to the reader, though also it follows from his background assumption of the descriptive adequacy of classical expected utility theory. Note first of all that if o_1 entails P, then the conjunction of P and o_1 is equivalent to o_1, so (Ramsey implicitly assumes) $\mathcal{D}es(o_1) = \mathcal{D}es(o_1 \,\&\, P)$. Thus, if $(o_1, P; o_2) \sim o_3$, where o_1 entails P and o_2 entails $\neg P$, then:

$$\mathcal{D}es((o_1, P; o_2)) = \mathcal{D}es(o_1).\mathcal{B}el(P) + \mathcal{D}es(o_2).(1 - \mathcal{B}el(P)) = \mathcal{D}es(o_3)$$

This is then rearranged to give us the definition of $\mathcal{B}el(P)$ as above. Importantly, because ratios of differences are preserved across positive linear transformations of $\mathcal{D}es$, $\mathcal{B}el(P)$ so-defined is *unique*. That is, there is only *one* function $\mathcal{B}el$ from $\boldsymbol{\mathcal{P}}$ into [0, 1] such that the above equalities hold. It follows that *to the extent that* $\mathcal{D}es$ accurately measures our subject's utilities, *and* that she values the relevant gambles according to the standard expected utility formula, *and* attaches the same value to o_1 and $(o_1 \,\&\, P)$ whenever $o_1 \vdash P$, *then* $\mathcal{B}el$ fixes the only way that her credences could possibly be (at least with respect to the propositions in $\boldsymbol{\mathcal{P}}$).

For future discussion, it is worth making Ramsey's implicit assumption explicit:

Indifference to Equivalent Conjunctions
For all P, Q, if $P \vdash Q$, then $P \sim (P \,\&\, Q)$

Ramsey also note two more conditions needed to ensure the coherence of his Definition 4. The first of these is that the value of $\mathcal{B}el(P)$ does not depend on the choice of outcomes and gambles satisfying the stated conditions. In effect, this is to place restrictions directly upon $\mathcal{B}el$ after it has been defined in terms of preferences. The second assumption is that for any gamble $(o_1, P; o_2)$, we will always be able to find some outcome o_3 such that $o_3 \sim (o_1, P; o_2)$. Making this latter assumption provides a very simple way of extending $\mathcal{D}es$, which in the first instance is only defined on $\mathcal{O}$, to $\mathcal{O} \cup \mathcal{G}$.

As a matter of fact, there is a further condition which Ramsey neglected to mention, which is needed to ensure that $\mathcal{B}el$ so-defined will always have a value within the [0, 1] interval:

For all $o_1, o_2 \in \mathcal{O}$ and $(o_1, P; o_2) \in \mathcal{G}$, if $o_1 \succcurlyeq o_2$, then $o_1 \succcurlyeq (o_1, P; o_2)$

Essentially, this condition states that a gamble is never strictly preferred to its best possible outcome. It is easy enough to see that without this condition—which is not implied by any of the foregoing axioms or assumptions—then Definition 4 may occasionally leave us with a value for $\mathcal{B}el(P)$ that is less than 0 or greater than 1.

It is interesting to note that at this stage, and without adding any further conditions on $\succcurlyeq$, $\mathcal{B}el$ need not be a probability function; indeed the only general properties which hold of it are:

(i) If $\top$ is necessary and $\bot$ is impossible, then $\mathcal{B}el(\top) = 1$ and $\mathcal{B}el(\bot) = 0$
(ii) $\mathcal{B}el(P) = 1 - \mathcal{B}el(\neg P)$
(iii) For some proposition(s) P in $\mathcal{P}$, $\mathcal{B}el(P) = 0.5$

There is nothing in the conditions stated so far that suggests that $\mathcal{B}el$ ought to be additive, nor even monotonic.[10] The reason for this flexibility is that the stated conditions place very few restrictions at all on gambles involving propositions which are *not* ethically neutral. It is entirely consistent with everything said so far that $P \vdash Q$, $o_1 \succ o_2$, and yet $(o_1, P; o_2) \succ (o_1, Q; o_2)$—in which case Definition 4 immediately implies that $\mathcal{B}el(P) \geq \mathcal{B}el(Q)$ and, hence, that $\mathcal{B}el$ is not a probability function.

Ramsey (1931, 180ff) goes on to define conditional probabilities using further conditions on preferences over more complicated gambles, and he argues that the resulting $\mathcal{B}el$ satisfies the laws of probability. I will not recapitulate that argument here: it is enough that Ramsey provides us conditions sufficient for pinning down a *unique* credence function, $\mathcal{B}el\colon \mathcal{P} \mapsto [0, 1]$, that supposedly represents the agent's credences—after all, it combines with the agent's utilities for outcomes to determine their preference ordering for two-outcome gambles in more or less the manner we pre-theoretically expect credence to do so. For our present purposes, it is incidental whether $\mathcal{B}el$ satisfies the conditions of the probability calculus, and the more general representation result requiring fewer preference conditions and a more flexible $\mathcal{B}el$ function is in many ways the more interesting of the two.

2. The problem of ethical neutrality

Despite its very early inception, there are several features that make Ramsey's system attractive, especially in comparison to later works. The theorems developed by von Neumann & Morgenstern and Anscombe & Aumann were in some respects a rediscovery of ideas already

[10] A credence function $\mathcal{B}el$ is monotonic iff, if $P \vdash Q$, then $\mathcal{B}el(P) \leq \mathcal{B}el(Q)$. Additivity implies monotonicity, but not *vice versa*.

present in 'Truth and Probability', but their appeal to extrinsically given probabilities limits their applicability, whereas Ramsey's system makes no such appeal. Savage's theorem was also founded on Ramseyan ideas, but Savage's system suffers from a number of defects not present in Ramsey's system. For instance, given the plausible assumption that Ramsey wanted to avoid impossible gambles (§2.1), the outcomes of a gamble are always consistent with the gamble's condition. Consequently, Ramsey's system seems to avoid anything like the constant acts problem that plagues Savage's system (see Joyce 1999, Chp. 3). Furthermore, the domain of Ramsey's $\mathcal{B}el$ is not limited to disjunctions of states, a feature of Savage's system which severely limits its overall usefulness (see Elliott MS-b). Another attractive feature of Ramsey's proposal is its strong uniqueness condition. We might contrast this with the monoset theorem of Jeffrey's, where the $<\mathcal{B}el, \mathcal{D}es>$ pair is only unique up to a fractional linear transformation.

All of this is achieved, however, on the basis of a highly problematic assumption about ethically neutral propositions, which I will now argue makes Ramsey's system untenable. My critical discussion of Ramsey's ideas focuses on this assumption as it raises unique problems not faced by the theorems I have considered in earlier chapters.

2.1 Why Ramsey needed ethical neutrality

Ramsey was right to reject Naïve Expected Utility Theory. If o is compatible with both P and $\neg P$, then it's entirely possible that the agent values $(o \,\&\, P)$ more (or less) than $(o \,\&\, \neg P)$. Any rational agent ought to take this into account when deliberating between gambles conditional on P with o as an outcome. For example, contrary to Naïve Expected Utility Theory, it's possible that the agent could be indifferent between $o_1 \sim o_2$ without thereby being indifferent between $(o_1, P; o_2)$ and $(o_2, P; o_1)$, if the truth or falsity of P makes a difference to how the agent values o_1 or o_2.

However, this point is conditional on o_1 and o_2 being each compatible with both P and $\neg P$. If instead we suppose that o_1 implies P, then $(o_1 \,\&\, P)$ is logically equivalent to o_1—and for Ramsey, if o_1 implies P, then the desirability of o_1 is just the desirability of $(o_1 \,\&\, P)$. Ramsey's characterisation of the $\mathcal{B}el$ function relies on this assumption. So, inasmuch as o_1 implies P and o_2 implies $\neg P$,

$$\begin{aligned}\mathcal{D}es((o_1, P; o_2)) &= \mathcal{D}es(o_1 \,\&\, P).\mathcal{B}el(P) + \mathcal{D}es(o_2 \,\&\, \neg P).(1 - \mathcal{B}el(P)) \\ &= \mathcal{D}es(o_1).\mathcal{B}el(P) + \mathcal{D}es(o_2).(1 - \mathcal{B}el(P))\end{aligned}$$

Note that this holds regardless of whether P is ethically neutral or not. In other words, if o_1 implies P and o_2 implies $\neg P$, and given Indifference to Equivalent Conjunctions, we can apply Naïve Expected Utility Theory to the gamble $(o_1, P; o_2)$.

Interestingly, Ramsey originally describes his outcome set $\mathcal{O}$ as a set of possible worlds, and it is part of Ramsey's background theory that every world individually determines the truth or falsity of any proposition. In particular, Ramsey assumed a broadly Wittgensteinian logical atomism—though he believed it possible to reformulate his theorem without these commitments (see his 1931, 177). We are to suppose that there exists a class of atomic propositions such that no two worlds are exactly identical with respect to the truth of these propositions, every atomic proposition can be true or false entirely independently of any others, and for every world w and atomic proposition P, there is another world w^* that differs only with respect to the truth of P. Every possible world on this picture is determined by the set of atomic propositions true at that world. Even setting aside the assumption of logical atomism, on an orthodox conception of propositions as sets of worlds, then for any given (determinate) proposition, a given world either is or is not a member of that proposition. Every world therefore determines either the truth or falsity of any proposition.

This leaves us with something of a puzzle: why did Ramsey alter his characterisation of the outcome set (as noted in §1.1)? It seems that if he limited his attention to gambles like (o_1, P; o_2), where o_1 implies P and o_2 implies $\neg P$, then he could have used preferences over *these* to define $=^d$ without needing to introduce the notion of ethical neutrality. The following piece of terminology will be helpful:

Definition 5: Impossible gambles

A gamble (o_1, P; o_2) is *impossible* iff P and $\neg P$ are consistent and either (o_1 & P) or (o_2 & $\neg P$) are inconsistent; (o_1, P; o_2) is *possible* otherwise

Where outcomes are possible worlds, every *possible* gamble $(o_1, P; o_2)$ conditional on a contingent proposition P must be such that o_1 implies P and o_2 implies $\neg P$. Where one of either P or $\neg P$ is impossible—say, $\neg P$—then the other must be necessary; in which case $\mathcal{B}el(\neg P) = 0$, $\mathcal{B}el(P) = 1$, and every o implies P, so $\mathcal{D}es(o) = \mathcal{D}es(o \,\&\, P)$. We can therefore *always* apply Naïve Expected Utility Theory to *possible* gambles, *if* the outcomes in $\mathcal{O}$ are worlds. So why did Ramsey not stick to his original characterisation of outcomes as worlds, and simply use preferences over possible gambles to define $=^d$?

The answer to this question can be discovered by considering again how Ramsey defines what it is for an agent to have a credence of ½ in a proposition. In particular, to determine whether P is of credence ½, we need to consider preferences over two gambles of the form (o_1, P; o_2) and $(o_2, P; o_1)$. The definition Ramsey gives us *only* makes sense if the outcomes o_1 and o_2 are *not* possible worlds. If o_1 and o_2 are possible worlds, then at least one of the two gambles is impossible, and if either gamble is impossible, then the reasoning behind the assignment of a credence value of ½ to the contingent proposition P is no longer valid.

Indeed, Ramsey recognised the difficulty here, and for this reason wrote that, at least for some outcomes o_1 and o_2 required for his definition, o_1 and o_2 "must be supposed so far undefined as to be compatible with both P and $\neg P$". Supposing for simplicity that P is atomic, we are presumably to take the outcomes o_1 and o_2 as *near-worlds*, which we can understand as propositions that are just shy of being maximally specific. Given his logical atomism, for every world w and every atomic proposition P, there is a proposition that *nearly* uniquely identifies w except for specifying whether P is true or not. In Ramsey's framework, a near-world with respect to an atomic proposition P is a disjunction of two worlds w^P and $w^{\neg P}$ that are identical with respect to all of their atomic propositions except for P.

The answer to our puzzle, then, is that Ramsey's set of outcomes cannot quite be the set of possible worlds *given* his strategy for defining $=^d$. For the pair of possible gambles $(o_1, P; o_2)$ and $(o_2, P; o_1)$ referred to in Definition 1, neither o_1 nor o_2 can imply either P or $\neg P$. It follows for the reasons given, then, that we cannot in general apply Naïve Expected Utility Theory to such gambles without appeal to ethically neutral propositions.

Before I move on to why the presumed existence of ethically neutral propositions is problematic, it is worth noting that Ramsey's **RAM1** seems to severely *understate* what he actually required for the adequacy of his measurement system. This is because, given how he proposed to define $=^d$, without changes elsewhere in his system Ramsey also required *either* that we have preferences over impossible gambles, *or* that every member of $\mathcal{O}$ is compatible with both the truth and falsity of some ethically neutral proposition. The argument for this proceeds by first noting that if $\mathcal{D}es(o_1) - \mathcal{D}es(o_2) = \mathcal{D}es(o_3) - \mathcal{D}es(o_4)$, then it should be the case that $(o_1, o_2) =^d (o_3, o_4)$. Suppose that $o_1 \sim o'_1$, so $\mathcal{D}es(o_1) - \mathcal{D}es(o'_1) = \mathcal{D}es(o'_1) - \mathcal{D}es(o_1)$. From **Definition 2**, we know that $(o_1, o'_1) =^d (o'_1, o_1)$ is only defined if the agent has preferences over some pair of gambles of the form $(o_1, P; o_1)$ and $(o'_1, P; o'_1)$ for some ethically neutral P of probability ½. It follows that either o_1 is compatible with P and $\neg P$, and likewise for o'_1, or at least one of these two gambles is impossible.

One might suppose that Ramsey was happy to deal with preferences over impossible gambles. This would have forced him to assume that there is an interesting difference between two impossible propositions $(o_1 \& P)$ and $(o_2 \& P)$, where both o_1 and o_2 entail $\neg P$ but $\neg(o_1 \sim o_2)$. For suppose that Ramsey had only one impossible proposition, $\emptyset$. Then $\mathcal{D}es(o_1 \& P) = \mathcal{D}es(o_2 \& P) = \mathcal{D}es(\emptyset)$, but $\mathcal{D}es(o_1) \neq \mathcal{D}es(o_2)$. For whatever value we take $\mathcal{D}es(\emptyset)$ to have, it is clear that this will lead to problems. For instance, suppose that $\mathcal{D}es(\emptyset) \neq \mathcal{D}es(o_1)$; o_1 and o_2 each imply P; o_3 implies $\neg P$; and $\mathcal{D}es(o_1) = x$, $\mathcal{D}es(o_2) = \mathcal{D}es(o_3) = y$. We require that $(o_1, o_2) =^d (o_1, o_3)$, for obviously $x - y = x - y$. However, the justification for the definition of $=^d$ in terms of preferences fails under these conditions:

$$(o_1, o_2) =^d (o_1, o_3) \text{ if and only if } (o_1, P; o_3) \sim (o_2, P; o_1)$$

But this holds just in case

$$\mathcal{D}es(o_1 \& P).\mathcal{B}el(P) + \mathcal{D}es(o_3 \& \neg P).(1 - \mathcal{B}el(P)) = \mathcal{D}es(o_2 \& P).\mathcal{B}el(P) + \mathcal{D}es(o_1 \& \neg P).(1 - \mathcal{B}el(P))$$

Supposing P is ethically neutral and is of probability ½, this equals

$$\tfrac{1}{2}x + \tfrac{1}{2}y = \tfrac{1}{2}y + \tfrac{1}{2}.\mathcal{D}es(\emptyset)$$

It follows that $=\mathcal{D}es(\emptyset) = x = \mathcal{D}es(o_1)$, which contradicts our initial assumption.

The only consistent way that Ramsey could have included impossible gambles in his system would have been to treat different impossible propositions as different objects of desire. Perhaps an appeal to impossible worlds would suffice for this purpose—the impossible prospect of being a married bachelor might be desired to a greater degree than the prospect of being a square circle. However, the move from worlds to near-worlds in his characterisation of the outcome set $\mathcal{O}$ strongly suggests that he desired to avoid impossible gambles.

And rightly so: restricting our attention to possible gambles seems the most plausible option. It is not obvious how we ought to treat preferences with respect to impossible propositions, if indeed there is more than one such proposition. For instance, it is implicit in Ramsey's system that if $o_1 \succcurlyeq o_2$, then $o_1 \succcurlyeq (o_1, P; o_2)$. As noted above, without some such assumption he is unable to show that the function *bel* is a credence function. However, if we know that $(o_1 \& P)$ is impossible, it is not obvious why this should be the case: It seems at least as plausible that o_2

$\succcurlyeq$ (o_1, P; o_2) in this case, as we know we are not going to receive o_1 in the event that P and there is only a potentially very small credence $\mathcal{B}el(\neg P) \leq 1$ of receiving o_2.

Thus, it looks as though Ramsey was implicitly assuming a stronger condition:

RAM1* For every $o \in \mathcal{O}$, there is at least one ethically neutral proposition P of credence ½ such that w is compatible P and $\neg P$

An interpretive difficulty arises here from the fact that **RAM1*** only makes sense if none of the outcomes in $\mathcal{O}$ are maximally specific, which is in conflict with Ramsey's original characterisation. Despite describing the elements of $\mathcal{O}$ as "totalities of events", the most charitable interpretation appears to be that Ramsey intended all his outcomes to be maximally specific *with respect to what the agent cares about*. Because agents do not care about the truth or falsity of ethically neutral propositions, any given o in $\mathcal{O}$ should be non-specific with regards the truth of any given ethically neutral proposition. In any case, as I will argue shortly, **RAM1** is already too strong an assumption. **RAM1*** is stronger still, and by a wide margin. Even where the former might be defended, the latter seems indefensible.

2.2 Problems with ethical neutrality

In looking at whether the notion of ethical neutrality is viable, we ought to start with Ramsey's own definition:

Definition 6: Ethical neutrality (Ramsey's original)
P is ethically neutral iff (i) if P is atomic, then $w^P \sim w^{\neg P}$, for all pairs of worlds w^P, $w^{\neg P}$ identical with respect to all their atomic propositions except for P, (ii) if P is non-atomic, then all of Ps atomic truth arguments are ethically neutral

So, an atomic proposition P is *ethically neutral* for an agent iff any two possible worlds differing in their atomic propositions only in the truth of P are always equally valued by that agent, and ethical neutrality for non-atomic propositions is understood in terms of atomic propositions. Ramsey here demonstrates commitment to another aspect of Wittgensteinian atomism: every non-atomic proposition can be constructed from atomic propositions using truth-functional connectives. We are able to locate such a proposition, if it exists, by considering the agent's preferences over worlds. As just noted, for some gambles ($o_1, P; o_2$) and ($o_2, P; o_1$), Ramsey requires that o_1 and o_2 are compatible with both P and $\neg P$. If we suppose for simplicity that P is atomic, then o_1 and o_2 are near-worlds with respect to P. It follows from Ramsey's definition then that (o_1 & P) ~ (o_1 & $\neg P$) and (o_2 & P) ~ (o_2 & $\neg P$). It does *not* yet follow that (o_1 & P) ~ (o_1) ~ (o_1 & $\neg P$), which Ramsey also required. However, we can take this as an unstated background assumption: if (o_1 & P) ~ (o_1 & $\neg P$), then (o_1 & P) ~ (o_1) ~ (o_1 & $\neg P$).

Sobel (1998, 241) has argued that there are few or no ethically neutral propositions in this sense. Consider the proposition *there are an even number of hairs on Dan Quayle*'*s head*. Sobel argues that this can be ethically neutral for 'almost no one':

> Though it is true that I do not care about Quayle's hair, there are worlds that differ regarding the truth of that proposition that, just because of that difference, differ in their values for me. I am

thinking of worlds in which I have bet money on this proposition! The argument … can be readdressed to atomic propositions, if such there be, to the conclusion that *no* atomic proposition is Ramsey-ethically-neutral for any of us. (1998, 248)

There seem to be two concerns here. The first appears to be something like the following: for any proposition whatsoever, we should be able to find a set of otherwise similar possible worlds where we have entered into a bet conditional on that proposition with desirable outcomes if things turn out one way, and undesirable outcomes if things turn out another way. Since we care about the outcomes of the bet, we will value the relevant worlds differently. However, this objection seems to have no hold given Ramsey's view: the relevant worlds are supposed to differ at the atomic level *only* with respect to the proposition in question. In all other respects—including, importantly, the payouts for any bets we may enter into—the worlds are supposed to be identical.

The second and more obvious worry is that Ramsey's conception of ethical neutrality requires the assumption of logical atomism for its cogency. Ramsey built his theory upon the assumption of logical atomism so that he could make sense of the idea of two worlds differing *only* with respect to a particular proposition. The notion is of little use to contemporary philosophers who by and large reject that aspect of Wittgenstein's view. If we are to give $\succcurlyeq$ a plausible interpretation *qua* preference relation, we had better not build our account of that relation's objects on a now-defunct account of propositions.

In his atomism-free reconstruction of Ramsey's theorem, Bradley (2001) supplies the following definition, intended to achieve the same purpose:[11]

Definition 7: Ethical neutrality (atom-free)
P is ethically neutral iff for all propositions Q (that are compatible with both P and $\neg P$), $(P \,\&\, Q) \sim Q \sim (\neg P \,\&\, Q)$

Tautological and impossible propositions will be trivially ethically neutral according to this definition. Clearly, however, we are interested only in non-trivially ethically neutral propositions. A common suggestion is that propositions such as *the tossed coin will land heads* constitute ethically neutral propositions of credence ½. Part of the reason why we use coin tosses occasionally when making decisions is because we have no intrinsic interest in whether the coin lands heads or tails. If Q is something like *there are dogs*, and P is *the tossed coin will land heads*, then it seems plausible that $(\mathrm{P} \,\&\, Q) \sim Q \sim (\neg \mathrm{P} \,\&\, Q)$.

However, there are strong reasons to think that *no* contingent propositions will be ethically neutral in the sense of Definition 7, for any minimally rational subject. Let P be *the tossed coin will land heads*, and take Q to be the proposition (*the tossed coin will land heads* & *I receive $100000*) *or* (*the tossed coin will not land heads* & *I get kicked in the shins*). Q is obviously compatible with both P and $\neg P$. However, $(\mathrm{P} \,\&\, Q)$ is equivalent to *the tossed coin will land heads* & *I receive $100000* while $(\neg \mathrm{P} \,\&\, Q)$ is equivalent to *tossed coin will not land heads* & *I*

[11] Bradley neglects to include the requirement that Q should be compatible with both P and $\neg P$, which leads to problems: letting Q be the necessary proposition $\top$, then P and $\neg P$ must be valued the same as $\top$; and letting Q be P, then P must have the same value as $(\neg P \,\&\, P)$ (i.e., $\neg\top$); so, by transitivity, all ethically neutral propositions must have the same value, which must be the value of both $\top$ and $\neg\top$. The final result is incompatible with, for instance, Jeffrey's decision theory, where $\top$ and $\neg\top$ are assumed to have different values.

get kicked in the shins. But for some very strange preference orderings, it's certainly not the case that (P & Q) ~ Q ~ (¬P & Q).

The point here generalises easily; there are no non-trivially ethically neutral propositions in this sense. Note that the issue here is not that no contingent proposition satisfies the definition *exactly*, while there may nevertheless be some propositions which *approximate* ethical neutrality. Rather, the upshot is that no proposition even comes *close* to satisfying the requirements of ethical neutrality. We will always be able to find countless many propositions Q that falsify the indifference requirements.

A refinement of Definition 7 might be useful. Instead of requiring (P & Q) ~ Q ~ (¬P & Q) for *all* Q compatible with both P and ¬P, Ramsey only requires the following:

Definition 8: Ethical neutrality (atom-free, refined)
P is ethically neutral iff o ~ (o & P) ~ (o & ¬P), for any outcome $o \in \mathcal{O}$ that is compatible with both P and ¬P

If there are no outcomes compatible with both P and ¬P, then P is trivially ethically neutral by this definition. Again, we can set such propositions aside; we are interested in non-trivially ethically neutral propositions. Definition 8 is weaker than Definition 7 because if Q is not in the outcome set $\mathcal{O}$, then there are no relevant gambles with Q as an outcome and we do not need to concern ourselves over whether (P & Q) ~ Q ~ (¬P & Q). More generally, if we assume that there are far fewer propositions in $\mathcal{O}$ than in $\mathcal{P}$, then the foregoing objection to Definition 7 is blocked. This will certainly be true if the outcomes in $\mathcal{O}$ are highly specific, as is the case in Ramsey's system.

With that said, it's still not obvious that any non-trivially ethically neutral propositions exist even in this weaker sense. Why should we suppose that there are *any* propositions P such that (non-trivially), o ~ (o & P) ~ (o & ¬P) for all $o \in \mathcal{O}$ compatible with P and ¬P? And moreover, if **RAM1*** is being assumed, why should we suppose that for *every* $o \in \mathcal{O}$, we will find such propositions? Without knowing the exact nature of the outcome space $\mathcal{O}$, we cannot even know whether there *are* any outcomes compatible with both P and ¬P, for an arbitrarily chosen proposition P. Ramsey explicitly stipulates that there must be at least one pair of outcomes compatible with *some* ethically neutral proposition of credence ½ and its negation—but this stipulation is meaningless inasmuch as we do not already know what proposition that may be. Unfortunately, Ramsey's discussion leaves the nature of $\mathcal{O}$ quite vague, making the matter impossible to judge.

We can circumvent this concern by stipulating that $\mathcal{O}$ contains, for each of a very wide range of propositions in $\mathcal{P}$, outcomes that are undefined with respect to that proposition. But even then, Ramsey gives us little reason to suppose that ethically neutral propositions exist relative to a given agent's preference ordering—still less that there are any such propositions that satisfy Definition 1. **RAM1** clearly cannot be defended as a condition of rationality, and it does not follow from Ramsey's background assumption of the descriptive adequacy of CEU. Ramsey's aim in the first instance was to develop a procedure for the measurement of credences, so unlike other intended uses for decision-theoretic representation theorems he did not require his conditions to be constraints of practical rationality; nevertheless, if his process is to be viable then it ought at least be *applicable*. It may not be impossible for a rational agent to satisfy the condition, but we still require good reasons to believe that most do—yet no reasons are forthcoming.

A related issue regards Ramsey's proto-functionalist attempt to define credences in terms of his measurement procedure: a definition of credences which relies centrally on a dubitable and unjustified existential assumption is of very limited interest for modern characterisational representationism. Are we to suppose that agents who falsify **RAM1** do not have credences? Ultimately, given his reliance upon ethically neutral propositions, Ramsey's system was not sufficient to establish the main upshot of 'Truth and Probability': that the laws of probability provide for us the logic of partial belief. Even if it is understood in terms of Definition 8, **RAM1** is a very shaky foundation for a measurement procedure, and still worse for a characterisation of credences.

Many expected utility representation theorems developed since 'Truth and Probability' have also made use of ethically neutral propositions, whether explicitly or implicitly. Davidson and Suppes (1956) develop a representation theorem similar to Ramsey's wherein they explicitly characterise and assume the existence of ethically neutral propositions. Others make implicit appeal to ethically neutral propositions, in the sense that they figure in the intended interpretation of the formal system, rather than being formalised directly. In this capacity, for instance, we find ethical neutrality in the theorem of Debreu (1959), where $\succcurlyeq$ is defined on pairs of outcomes, which are understood as representing two-outcome gambles conditional on some ethically neutral P for which the agent has a credence of ½. Fishburne (1967) makes implicit appeal to ethically neutral propositions of credence ½ along very similar lines. Each of these works appear to require an understanding of ethical neutrality in something like the senses of Definition 7 or Definition 8 (each for essentially the same reason that Ramsey required the notion), and thus they inherit the problems associated with his use of ethically neutral propositions.

References

Adams, E. W. (1975). *The Logic of Conditionals*. Reidel.

Bradley, R. (1998). 'A Representation Theorem for a Decision Theory with Conditionals.' *Synthese* **116** (2): 187-229.

—— (2001). 'Ramsey and the Measurement of Belief'. In *Foundations of Bayesianism*. D. Corfield and J. Williamson, Eds. Kluwer Academic Publishers: 263-90.

Davidson, D. and Suppes, P. (1956). 'A finitistic axiomatization of subjective probability and utility.' *Econometrica* **24** (3): 264-75.

Debreu, G. (1959). 'Cardinal utility for even-chance mixtures of pairs of sure prospects.' *The Review of Economic Studies* **28** (3): 174-7.

Elliott, E. (forthcoming). 'Ramsey without Ethical Neutrality: A New Representation Theorem.' *Mind*.

—— (MS-a). 'A Representation Theorem for Frequently Irrational Agents.'

—— (MS-b). *Probabilism, Representation Theorems, and Whether Deliberation Crowds out Prediction*. Unpublished manuscript. Retrieved from https://drive.google.com/file/d/0BzcNaqBmo6btc1NOVm9KZTlpYWs/view.

Fishburn, P. C. (1967). 'Preference-based definitions of subjective probability.' *The Annals of Mathematical Statistics* **38** (6): 1605-17.

Jeffrey, R. C. (1983). 'Bayesianism with a human face.' *Testing Scientific Theories, Minnesota Studies in the Philosophy of Science* **10**: 133-56.

Joyce, J. M. (1999). *The foundations of causal decision theory*. Cambridge University Press.

Ramsey, F. P. (1931). 'Truth and probability'. In *The Foundations of Mathematics and Other Logical Essays*. R. B. Braithwaite, Ed. Routledge: 156-98.

Sobel, J. H. (1998). 'Ramsey's Foundations Extended to Desirabilities.' *Theory and Decision* **44** (3): 231-78.