HOW TO CO-EXIST WITH NONEXISTENT EXPECTATIONS

RANDALL G. MCCUTCHEON

Abstract. We address the problem that gambles having undefined expectation pose for decision theory. Observing that to place a value on such a gamble exposes one to a finitary diachronic Dutch Book, we defend a variant of Mark Colyvan's "Relative Expected Utility Theory" (**REUT**), noting that it has the property of never preferring a gamble X to an identically distributed gamble Y. We demonstrate, however, that even **REUT** subscribers succomb to diachronic incoherence should they assign infinite expectation to a gamble they actually confront. In a final section, we use basic principles of anthropic reasoning (as formulated by Brandon Carter) to show why one needn't ever do so.

1. A Diachronic Dutch Book against Pasadena Game Idealists

In Harris Nover and Alan Hájek's *Pasadena Game* (2004) you win X dollars, where

$$P\Big(X = \frac{(-1)^{n-1}2^n}{n}\Big) = 2^{-n}, \quad n = 1, 2, \ldots .$$

Since the expectation series for X,

$$\sum_{r\in\mathbf{R}} rP(X = r) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}2^n}{n} \cdot 2^{-n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

converges conditionally (to $\log 2$), X has no defined expectation.[1] Expected utility theory is, therefore, silent concerning the value of this game.[2] Nover and Hájek however write:

> It is an uncomfortable silence. For intuition tells us...that we can make meaningful comparisons between the Pasadena game and other games. It is clearly worse than the St. Petersburg game [$P(X = 2^n) = 2^{-n}$], for starters. It is clearly worse than a neighbouring variant of the game –call it the Altadena game–in which every pay-off is raised by a dollar. (...) And the Pasadena game is clearly better than a 'negative' St. Petersburg game, in which all the pay-offs of the St. Petersburg game are switched in sign. Yet expected utility theory can say none of this.

It does seem that one cannot coherently value the Pasadena gamble relative to the null gamble. For suppose that you value it at $\log 2$. (The problem will arise

[1] Like any conditionally convergent series, the expectation series can thus be made to diverge (or converge to any finite value whatsoever) by rearrangement of its terms. On the other hand, as noted by Kenny Easwaran (2008), X does have a weak expectation of $\log 2$. Since weak expectations are invariant under rearrangements, $\log 2$ therefore has some claim to be the presumptive value of the game, if it has one.

[2] We take utility to be linear with respect to currency, and in particular unbounded.

for any finite value.) Then you can be Dutch Booked as follows. First you are offered the chance to pay $.69 < \log 2$ to play the game. You accept and are put to sleep. While you sleep, your payoff will be determined as follows. First, a fair 4-sided die is rolled repeatedly until something less than a 4 is rolled. Let N be the number of rolls it takes. Repeat the procedure and let M be the number of rolls it takes the second time. Finally let Y be the result of a rolling of a fair 3-sided die. If $Y = 3$, let $n = 2M$. Otherwise, let $n = 2N - 1$. Your payoff is now $\frac{(-1)^{n-1}2^n}{n}$ dollars (a Pasadena gamble). Note: if $Y < 3$ and $n = 2N - 1$ you are winning money, whereas if $Y = 3$ and $n = 2M$ you are losing money.

When all of the above is completed, you are awakened and told the value N (but not M and not Y). So you don't know whether or not you've won, nor how much you've lost if you've lost. You do however know that if you've won, you've won 2^{2N-1} dollars. At this point, you are given the opportunity to annul the gamble (but you don't get your .69 back). As the expectation series for your payoff at this stage is

$$\frac{2}{3} \cdot 2^{2N-1} + \sum_{M=1}^{\infty} \frac{1}{2^{2M}} \cdot \frac{-2^{2M}}{2M} = \frac{2}{3} \cdot 2^{2N-1} - \frac{1}{2} - \frac{1}{4} - \frac{1}{6} - \cdots = -\infty,$$

you are compelled to accept this offer (indeed, you were let off easy–you would have paid any finite amount to annul the gamble). You have thus lost a sure .69.

The upshot is that if you assign faithful implementations of Pasadena gambles values and believe you are being offered one then your decision theory is diachronically incoherent in a strong sense; you are vulnerable to a finitary Dutch Book.[3]

2. Other responses

Hájek and Nover (2006) write:

> Here are...possible responses to the Pasadena game (...): 1. The game is coherent, and decision theory cannot handle it–too bad for decision theory. Compare: Russell's paradox was a decisive blow against Frege's set theory. Too bad for that set theory. 2. The game is incoherent, so it is not a black mark that decision theory cannot handle it–too bad for the game. Compare: the town barber, who shaves all and only those in the town who do not shave themselves, poses no problem for logic, or for anything else; he simply cannot exist, because the specification of him is incoherent. Too bad for the barber. (...) We maintain response (1)....

Hájek and Nover's appeal to Russell's paradox in (1) is problematic. The challenge for set theory upon discovery of the "Russell object" was to sidestep paradox while retaining such sets as the sincere mathematician requires. The standard

[3] By *Finitary Dutch Book* we intend a finite sequence $X_1, \cdots X_N$ of gambles, either offered all at once (a *synchronic* DB) or across time, with information gathering conducted between bets (a *diachronic* DB; in the diachronic case we allow that N, so long as $E(N) < \infty$, and the X_i may be random variables) deemed individually favorable but together entailing an almost sure net loss and having the property that $\sum_i E(\min\{X_i, 0\}) > -\infty$ almost surely.

response to the paradox, accordingly, has been to build into set theory appropriate admissibility criteria on sets. If we likewise view the Pasadena game's challenge to decision theory as that of sidestepping paradox while retaining such gambles as the sincere decision theoretic agent requires, then the best response to the game should be (just what Hájek and Nover don't want) to build into decision theory appropriate admissibility criteria on comparisons.

On the topic of restrictions, Hájek and Nover write: "There are two possible lines of attack–neither satisfactory, in our opinion. (...) *Restrict decision theory to finite state spaces* (...) *Restrict decision theory to bounded utility functions.*" Nicholas J.J. Smith (2014), notably, defends a hybrid of these:

Rationally negligible probabilities (RNP): For any lottery featuring in any decision problem faced by any agent, there is an $\epsilon > 0$ such that the agent need not consider outcomes of that lottery of probability less than ϵ in coming to a fully rational decision.

For variables of finite expectation, **RNP** is fairly harmless; as $\epsilon \to 0$, the relative effect of employing **RNP** instead of standard expectation tends to zero with ϵ. In this sense, **RNP** is an extension of (finite) expected utility theory. This observation can be used to respond to Hájek (2014), who seeks to discredit **RNP** using a zero expectation gamble (credited to John Matthewson) which **RNP**'s vanishing error term causes to look favorable (if infinitesimally). But although finite expectation gambles fail to discredit this more rebust limit version of **RNP**, even it sanctions some embarrassing preferences.

Game: A dime is tossed until it comes up heads (on the nth toss). Then a nickel is tossed. If the nickel comes up heads, you win 2^n dollars. If it comes up tails, you lose 2^n dollars.

RNP instructs us to ignore outcomes having probability below some threshold $\epsilon > 0$. So there is an n depending on ϵ such that **Game's** value is:

$$(.25)(2)-(.25)(2)+(.125)(4)+(-.125)(4)+\cdots+(2^{-n-1})(2^n)+(-2^{-n-1})(2^n)=0.$$

Thus the value of **Game** is zero, independently of ϵ.

Now consider a variant of **Game** in which, if the nickel lands heads, a penny is also tossed. If the penny lands heads, it is added to your winnings. Otherwise everything is as before. This variant's value is:

$$\begin{aligned}&(-.25)(2)+(.125)(2.01)+(.125)(2)+(-.125)(4)+(.0625)(4.01)+(.0625)(4)\\&\quad+(.0625)(8)+\cdots+(2^{-n-1})(2^{n-1}+.01)+(2^{-n-1})(2^{n-1})+(2^{-n-1})(2^n)<-.49.\end{aligned}$$

This inequality holds independently of ϵ.

So according to **RNP**, **Game** has value $V_1 = 0$ and the variant $V_2 < -.49$, independently of ϵ. But the only difference between **Game** and the variant is that in the latter there is a penny that you might get to keep.

3. Relative Expected Utility Theory

The above failure notwithstanding, there *is* a good way to avoid paradox by restriction; one can simply hold that two gambles should be deemed comparable if and only if their difference has a defined (finite or infinite) expectation. The Pasadena gamble does not have a defined expectation, so is on this view incommensurable with the null gamble. That it has a coherent definition is neither here nor there...what seals the case is that it cannot coherently be assigned a value.

Mark Colyvan has championed an approach that has been converging to the above. First he pointed out (see Colyvan 2006) that a preference for Altadena over Pasadena can be established by dominance reasoning. Next (see Colyvan 2008) he formulated a "relative expected utility theory", a joint extension of finite expected utility theory and just this sort of dominance reasoning. Finally in Colyvan and Hájek (2016), an emended version of this theory was put forth.[4]

Relative Expectation Utility Theory (REUT): Suppose gamble A pays a_i and gamble B pays b_i in state S_i, $i \in \mathbf{N}$, and put $p_i = Prob(S_i)$. Then A is preferable to B when the expectation of $A - B$ is defined and positive; that is, when $E(A - B) > 0$. (In particular $E\big(\min\{A - B, 0\}\big)$ must be finite, though $E\big(\max\{A - B, 0\}\big)$ needn't be.) Equivalently, if $REU(A, B) = \sum_{i \in \mathbf{N}} p_i(a_i - b_i)$ is invariant under rearrangment of indices and positive.

REUT recovers the "clear" comparisons (Altadena preferable to Pasadena, etc.) cited by Nover and Hájek. A further important (and not, it seems, so obvious) virtue is that it never indicates a preference for a gamble X over an identically distributed gamble Y. (We prove this in an as-yet unpublished manuscript.)

4. Objections to **REUT**

As mentioned, agents subscribing to **REUT** do not ever prefer a gamble X to an identically distributed gamble Y. They are also not subject to conventional (finitary) Dutch Books. These are good properties for a decision theory to have, but there are other desiderata that **REUT** seemingly fares not-so-well on.

First Objection: Group Dutch Books

Utility sharing groups of agents subscribing to **REUT** are vulnerable to "Group Dutch Books" if the agents comprising the group can make unilateral decisions and believe that it's possible to confer a good of unbounded expected utility:

San Marino Game: Stanley and Stella are **REUT** subscribers married in the state of Louisiana, where they have what is known as the Napoleonic Code (according to which what belongs to the wife belongs to the husband also and vice versa). Stanley (together with a lawyer acquaintance) has devised a plan capitalizing on the fact that a gift of money on Stella's birthday is theoretically free under the Code. To liven things up, he presents to Stella a Huntington Library

[4] Our formulation isn't quite the same as theirs. In particular, of their two emendations we've removed one and altered another; it's clear, however, that this is the one Colyvan sought.

postcard with an enclosed coupon reading "Happy Birthday Stell. Luck is believing you're lucky! This coupon good for one Pasadena gamble, payable in dollars." Stella complains to her sister (Blanche) that although she has accepted the "gift" she's realized she could end up owing Stanley money under its terms. Blanche (a sometime adjunct scholar) sees an opportunity and offers to administrate the gamble. First, however, Blanche shows Stella a partition $\{P_i : i = 1, 2, \ldots\}$ of the naturals such that, for every i, the expectation of the Pasadena gamble in question exists and is equal to $-\infty$ conditional on the gamble paying from a state $n \in P_i$, and shows Stanley a partition $\{Q_j : i = 1, 2, \ldots\}$ of the naturals such that, for every j, the expectation of the Pasadena gamble in question exists and is equal to $+\infty$ conditional on the gamble paying from a state $n \in Q_j$. She explains to them both that the Pasadena gamble pays P_n with probability p_n, $n = 1, 2, \ldots$ and that they will learn which cell from their own partition contains n and be given a chance to cancel their position (for a price) after receiving this information but before learning the value of n. Blanche now puts Stanley and Stella to sleep in separate rooms. While they are asleep she rolls dice to determine n, then wakes them up. Blanche now goes to Stella and tells her the unique value i for which $n \in P_i$. At this point Stella realizes that the expected value of the gamble, from her perspective, is $-\infty$. Blanche now offers, as she promised, to sell her a short position in the same gamble–for a mere \$5. Stella gives Blanche the money, effectively cancelling her long position. Blanche then goes to Stanley and tells him the unique value j for which $n \in Q_j$. At this point Stanley realizes that the expected value of the gamble, from his perspective, is $+\infty$. Blanche now offers, as she promised, to sell him a long position in the same gamble–for \$150. He agrees to the transaction, which effectively cancels both his own short position and her long. The Stanley/Stella team has lost a sure \$155–a "Group Dutch Book".

Second Objection: Infinitary Dutch Books

A second objection is that single agents are subject to *infinitary* Dutch Books under **REUT**. **REUT** does not assume that utility functions are bounded or that state spaces are finite. But Vann McGee (1999) used an "airtight Dutch Book" to seemingly show that the combination of infinite state space and unbounded utility function leads expected utility theory subscribers to decision theoretic incoherence. McGee constructed an infinitary Dutch Book with payoffs w_i, $i = 1, 2, \ldots$.

What McGee failed to flag, however, was that his Dutch Book is infinitary, i.e. $\sum_i E\big(\min\{w_i, 0\}\big) = -\infty$. So while expected utility theory sanctions each bet considered by itself, simultaneous acceptance of them would appear to violate the spirit of **REUT**. Indeed, it's an easy matter to express any given infinite expectation wager as an infinite series of finite expectation wagers, so clearly **REUT** must be taken to implicitly sanction against simultaneous acceptance (or acceptance within any bounded window of time) terms constituting such a series.

But that won't quite do. If you believe your lifespan to be finite almost surely but to have infinite expected duration, a McGee-style Dutch Book can be administered to you across time in such a way that the quantity of utility you risk in any fixed unit of time (a day, say) is universally bounded above. To see how, consider the following twist on an experiment from Arntzenius, Elga and Hawthorne (2004):

Trumped. Donald Trump has just arrived in Purgatory. God explains that the duration X of his afterlife will be an instance of the St. Petersburg random variable (equal to 2^n with probability 2^{-n}), in days. Variety is the spice of the afterlife, however, and Trump will have the option, on Day 1, of spending that day in Hell in exchange for Days $4, 8, 12, \dots, 1020$ in Heaven (each contingent on his being around). This is an expected two days and he finds Heaven to be as pleasant as Hell is unpleasant, so he takes the deal. On Day 2 he agrees to spend that day in Hell in exchange for Days $1024, 1028, 1032, \dots, 2^{18}-4$ in Heaven (again an expected two days). And so forth...each day numbered $4n$ Trump spends in Heaven, but he spends the other days in Hell in exchange for contingent days $2^{8k+2}, 2^{8k+2}+4, \dots, 2^{8k+10}-4$ in Heaven, $k = 2, 3, \dots$. Expected utility theory recommends each bet, but the result of accepting them all is that Trump spends at least three-fourths of his afterlife in Hell.

Though Trump only accepts one wager per day, though their negative payoffs are bounded below and though there are almost surely only finitely many such wagers, the door is left open to the Dutch Bookie by the fact that their negative expected payoffs nevertheless sum to minus infinity, due to the fact that the expected number of wagers is infinite.

This sort of disaster can't befall Trump if he knows his afterlife to have expected duration (in days) $E(X) = L < \infty$. To see this, suppose Trump values x days in Heaven at x utils, x days in Hell at $-x$ utils, and x days in Purgatory at 0 utils. We will show that if Trump makes arbitrarily many gambles concerning his daily whereabouts that are deemed fair by expected utility theory, then the expectation of $B =$ (total days spent in Heaven - total days spent in Hell) is zero.

To begin, note that

$$E(X) = \sum_{n=1}^{\infty} P(X \geq n) = L.$$

From the convergence of this series, it follows that $NP(X \geq N)$, which is at most twice the value of $\sum_{n=\lfloor \frac{N}{2} \rfloor}^{N} P(X \geq n)$, tends to zero as $N \to \infty$. For $n \in \mathbf{N}$ let $p_n = \frac{P(X \geq n)}{L}$. Then $\sum_{n=1}^{\infty} p_n = 1$. Note that p_n is the expected density of nth days when Trump's afterlife is iterated without end. Observe also that

$$P(X = n) = \frac{p_n - p_{n+1}}{p_1}.$$

All of Trump's fair gambles may be subsumed into a single gamble expressed by a sequence $(x_n)_{n=1}^{\infty}$ taking values in $[-1, 1]$, where $x_n = -1$ indicates Trump spends Day n in Hell, $x_n = \frac{1}{2}$ indicates Trump spends half of Day n in Heaven, etc. By fairness, $\sum_{n=1}^{\infty} p_n x_n = 0$. Note that

$$\big|p_{N+1} \sum_{n=1}^{N} x_n\big| \leq (N+1)p_{N+1} \leq (N+1)P(X \geq N+1) \to 0,$$

so that

$$
\begin{aligned}
E(B) &= \sum_{n=1}^{\infty} P(X=n) \sum_{i=1}^{n} x_i \\
&= \sum_{n=1}^{\infty} \frac{p_n - p_{n+1}}{p_1} \sum_{i=1}^{n} x_i \\
&= \frac{1}{p_1} \lim_{N\to\infty} \Big((p_1-p_2)x_1 + (p_2-p_3)(x_1+x_2) + \cdots (p_N - p_{N+1})(x_1+\cdots+x_N) \Big) \\
&= \frac{1}{p_1} \lim_{N\to\infty} \Big(\sum_{n=1}^{N} x_n p_n + -p_{N+1} \sum_{n=1}^{N} x_n \Big) = 0.
\end{aligned}
$$

Third Objection: Missed Arbitrages

Adam Elga (2010) has an arbitrage argument against imprecise credences that can be turned against the **REUT** subscriber. Suppose you are offered a dollar to take a long position in X, a no-expectation random variable. If you subscribe to **REUT** then X and the dollar are not commensurable, so if you believe that the offer is made in good faith then you'll presumably decline it. Moments later you are offered a dollar to take a short position in X. Again you decline. Nothing changes if we assume that you have prior knowledge of the protocol. That's irrational, as accepting both offers strictly dominates rejecting them.

5. The Nature of **REUT**'s Idealization

To help motivate our response to the difficulties raised in the previous section, consider the following passage, from McGee (1999):

> ...a global plan cannot afford to ignore exotic possibilities, or to fail to allow for unusually complicated systems of acts and consequences. Now no one would hope for a plan that would invariably enable us to overcome adverse circumstances or vicious enemies, but one would at least hope for a plan that would enable us to avoid being defeated by our own ill-planned actions.

We do not deny that our credence functions cannot afford to ignore exotic possibilities. (Recall the failures of **RNP** in Section 2.) For fixed n, Trump for example ought not to assign zero probability to the prospect of his having an afterlife 2^n days in duration. That hardly implies, however, that he should assign this prospect the face value probability 2^{-n}! To ignore an exotic possiblity is to assign it zero credence, so there are many ways to "not ignore" such possibilities that fall far short of taking their advertised chances at face value. In particular, one can assign such events non-zero probabilties in such a way that the expectations of decision-theoretic quantities (the duration of Trump's afterlife, for example) come out finite. And, as we have seen, there is good reason for doing so. Namely, so that one isn't defeated by one's own "ill-planned actions".[5]

Note however that in doing so our **REUT**-subscribing agent has encountered a problem. Her only advantage over an agent subscribing to conventional expected utility theory was her ability to make comparisons between some pairs of no-expectation gambles. She accomplished this in a way that avoids finitary Dutch Books and never leads her to prefer a gamble X to an identically distributed gamble Y, but it now seems that in order to avoid Group Dutch Books, Infinitary Dutch Books and Missed Arbitrages, she now needs to shun credence functions according to which real decision-theoretic quantities (such as her expected lifespan) come out as having infinite expectation. But having done this, she can no longer encounter gambles that, by her own lights, have no expectation! Has she then eradicated her only advantage over the expected utility theoretician?

We don't think she has, though our reasons are subtle, and possibly contentious; we defer their presentation to the final section of the paper. Even for those who do insist that **REUT** is a formally vacuous extension of EUT, however, it may provide a means to simplify greatly many decision problems. Consider for example the following trivial theorem, which shows that as doubts about veracity vanish pointwise, correct action converges to **REUT**'s dictates.

Theorem. Suppose that a no-expectation gamble A pays a_i and a different no-expectation gamble B pays b_i in state S_i, where $Prob(S_i) = p_i$, $i = 1, 2, \ldots$.

a. **REUT** favors A to B if and only if there is a $\delta > 0$ having the property that for every finite set $F \subset \mathbf{N}$ satisfying $\sum_{n \in F^c} p_n < \delta$,

$$\sum_{i \in F} p_i(a_i - b_i) > \delta.$$

b. **REUT** favors A to B if and only if there is a $\delta > 0$ having the property that for any sequence $(x_n)_{n=1}^{\infty}$ with $0 \leq x_n \leq 1$ satisfying $\sum_{n \in \mathbf{N}} x_n(|a_n| + |b_n|) < \infty$ and $\sum_{n \in \mathbf{N}}(1 - x_n)p_n < \delta$,

$$\sum_{i \in F} x_i p_i(a_i - b_i) > \delta.$$

In b. one should think of x_n as the expected ratio of relative delivered to relative promised utility conditional on state S_n. (If $a_n - b_n = 10$ and $x_n = .7$ then the expected gain from $A - B$ in state n is 7.) The restriction $\sum_{n \in \mathbf{N}} x_n(|a_n| + |b_n|) < \infty$ is not technically necessary, though it will be satisfied for any agent who believes that the gambles being advertised as A and B are actually finite in expectation.

The content of the theorem, then, is that **REUT** favors A to B precisely when any sufficiently faithful joint implementation of the gambles consistent with restriction to finite expectation results in an approximation of A that is preferable to the

[5] One may protest that the proposition stating that, for every n, the "actual chance" is indeed 2^{-n} that Trump will have an afterlife 2^n days in duration, is itself an "exotic possibility" that we should not ignore. We address this issue in the next section.

corresponding approximation of B. The relationship of **REUT** to comparisons between commensurable infinite or no expectation (on their face) gambles, then, is like that of EUT to comparisons between finite expectation gambles; it makes the correct recommendation whenever the implementations are faithful enough.

6. Self indication and infinite expectation

In this final section we address the question of whether banishing variables of infinite expectation is in conflict with "regularity" assumptions on which one should not assign probability zero to any theory (e.g. *offers advertised as St. Petersburg gambles are genuine with positive probability*) that is not logically contradictory. In particular, we consider the following passage from Hájek (2006):

> Suppose I offer you the St. Petersburg game. You don't believe me; in fact you assign probability one in a trillion to the offer being genuine. Still, the paradox has a hold on you: for now the expectation of the game is a trillionth of infinity, which is still infinity.

But for this point to be valid the agent (Trump, say) would have to assign positive real probability to veracity *after* applying any relevant "anthropic reasoning", and varieties of anthropic reasoning on which such paradoxes remain uncontentiously live are problematic. We believe that the correct method is that of Brandon Carter (1983; see also A. Lewis 2001), who writes:

> "In a typical application of the anthropic (self-selection) principle, one is engaged in a scientific discrimination process of the usual kind in which one wishes to compare the plausibility of a set of alternative hypotheses, $H(T_i)$, say, to the effect that respectively one or other of a corresponding set of theories $T_1, T_2, \ldots$ is valid for some particular application in the light of some observational or experimental evidence E, say. Such a situation can be analysed in a traditional Bayesian framework by attributing *a priori* and *a posteriori* plausibility values (i.e. formal probability measures), denoted by p_E and p_S, say, to each hypothesis respectively before and after the evidence E is taken into account, so that for any particular result X one has
>
> $$p_E(X) = p_S(X|E),$$
>
> the standard symbol $|$ indicating conditionality. According to the usual Bayesian formula, the relative plausibility of two theories A and B, say, is modified by a factor equal to the ratio of the corresponding conditional *a priori* probabilities $p_S(E|A)$ and $p_S(E|B)$ for the occurrence of the result E in the theories, i.e.
>
> $$\frac{p_E(A)}{p_E(B)} = \frac{p_S(E|A)}{p_S(E|B)} \frac{p_S(A)}{p_S(B)}.\text{"} \tag{1}$$

The "Selected" or "Subjective" probability function p_S in (1) is related to an "Original" or "Objective" probability function p_O by $p_S(\cdot) = p_O(\cdot|S)$: "$S$ denotes...the selection conditions that are implied by the hypothesis of application of the theory to a concrete experimental or observational situation, but which are not necessarily included in the abstract theory" on which p_O is based.

It's implicit from Carter's own usage of the principle that a "theory" meanwhile is something like a measure on the set of universe histories rather than a chance or ineliminably indexical event such as "this toss of this coin lands *heads*". Examples of "theories" from Carter (1983) include the hypotheses: life is very rare, even in favorable conditions; gravitational coupling strength is fixed across time; and, the expected average time $\bar{t}$ intrisically most likely for the evolution of a system of observers intelligent enough to comprise a scientific civilization such as our own is geometrically small relative to the main sequence lifetime τ of a typical star.

Indeed, where A is a theory your nomologically accessible evidential counterparts (possible beings with thoughts indiscriminable from your own) should intend by "A" the exact same proposition you intend by "A"; in particular, the two utterances should be associated with the same truth value. (Your counterparts aren't actual beings contemplating a different coin or counterfactual beings for whom the coin landed otherwise than it actually did.) The importance of this restriction cannot be overstated, for Carter wishes to employ the identity

$$\frac{p_S(A)}{p_S(B)} = \frac{p_O(A)}{p_O(B)}, \tag{2}$$

which will not in general be valid for "non-theory" events A and B.[6]

So long as one assigns positive probability to the theory that there are *no* infinite expectation gambles, the expectation one ought to assign to an encountered St. Petersburg gamble after applying Carter's anthropic principle isn't infinite, but finite. To illustrate, suppose that Trump, on his first (and only) day in Purgatory, is offered a St. Petersburg variable X of days in Heaven in exchange for Y days in Hell, where Y is either 10 or 200 based on the toss of a fair coin. (Trump is told the value of Y prior to making his choice.) We again suppose that Trump values x days in Heaven at $+x$, x days in Hell at $-x$, and x days in Purgatory at 0. If Trump refuses the offer, his afterlife terminates at the end of the day. If he accepts, it terminates upon settlement. Trump is typical; every other conscious being in the universe has the same afterlife experience (we assume their pre-afterlife lifetimes are negligible) and that they are all rational and know the relevant protocols.

Suppose further that Trump entertains exactly two competing theories about the universe. Theory A says that the St. Petersburg offers X that one encounters in the afterlife are genuine. Theory B says that they are not genuine; in fact, their true expectations are precisely 50 days. Prior to reasoning anthropically, he is indifferent beween these two theories. Finally, assume for emphasis that Trump has amnesia upon awakening each morning, so that he awakens to find himself in the afterlife but uncertain as to whether it's Heaven, Hell or Purgatory–it's in this state that he assigns Theory A credence $\frac{1}{2}$. According to the sort of reasoning

[6] Cf. Sleeping Beauty, where the majority intuition is that (2) fails for $A = heads$, $B = tails$, vs. Bostrom's "Presumptuous Philosopher", where the majority intuition is that (2) holds for $A =$ *trillion trillion persons* and $B =$ *trillion trillion trillion persons*. See Bostrom (2007).

Hájek implicitly cites, then, the expected number of days Trump will spend in Heaven, should he accept, is $(\frac{1}{2})\infty + (\frac{1}{2})50 = \infty$. He should therefore accept, regardless of the value Y.

If we take Carter's anthropic principle into account, however, this computation breaks down. Assume for the moment that every being accepts the offer of X days in Heaven for Y days in Hell when $Y = 10$, but refuses when $Y = 200$. Let E be the event "I am now in Purgatory". Conditional on Theory B, half of all agents encounter $Y = 10$ and accept the ensuing offer; these agents spend an average of 61 days in the afterlife (1 in Purgatory, 10 in Hell, and an expected 50 in Heaven). The other half encounter $Y = 200$ and refuse the ensuing offer; these agents spend 1 day in the afterlife (in Purgatory). It follows that the expectation of Trump's afterlife conditional on B is 31 days, so that $p_S(E|B)$ is the multiplicative inverse of this expectation, i.e. 1/31.

Next assume, for the time being, that $E(X|A)$ is large but finite. Then $L = 6 + \frac{1}{2}E(X|A)$ is the expectation of Trump's afterlife conditional on A, so that $p_S(E|A)$ is equal to L^{-1}. One therefore has

$$\frac{p_E(A)}{p_E(B)} = \frac{p_S(E|A)}{p_S(E|B)}\frac{p_S(A)}{p_S(B)} = \frac{L^{-1}}{1/31}\frac{(1/2)}{(1/2)},$$

from which it follows that $p_E(A) \approx 31L^{-1}$ and $p_E(B) \approx 1$. Trump now computes the expectation of X as follows:

$$(3) \quad E(X) = p_S(A)E(X|A) + p_S(B)E(X|B) \approx 31L^{-1}(2L - 12) + 1(50) \approx 112.$$

Since the error in this approximation tends to zero as $E(X|A)$ increases, there is a compelling case for asserting that, in fact, $E(X) = 112 < 200$ when $E(X|A) = \infty$.[7] This expectation would be still lower should the agents accept when $Y = 200$; therefore (since we are assuming they are rational) they do not.

If that's right, then Hájek wasn't entitled to the claim that assigning even a tiny positive probility to a claimed St. Petersburg variable's veracity requires you to assign the variable an infinite unconditional expectation. Indeed, Carter's anthropic principle establishes something like the opposite. Namely, that assigning positive probability to the theory that faithful St. Petersburg gambles are impossible requires you to assign gambles advertised as St. Petersburg gambles finite unconditional expectation whenever they are encountered.

This computation dissolves, in particular, the worries of Section 4 without making **REUT** a vacuous extension of EUT. Note however that the computation cannot be adapted to arrive at a (unique) definite expectation for the Pasadena gamble; one would (in theory) obtain finite upper and lower bounds on the expected utility of Pasadena gamble by this method, but where in the resulting interval one landed

[7] One might even attempt to endow the calculation (3) with formal validity in the infinite expectation case by letting L be an appropriate infinite hyperreal (so that L^{-1} is infinitesimal). We shall leave this pursuit to others, however.

by any proposed limiting procedure would be sensitive to the order of the indices. This reassures one that **REUT** is on the right track. It also supports Nover and Hájek's (2004) contention that "the Pasadena game is more paradoxical than the St. Petersburg game in several respects."

References

Arntzenius, Frank, Adam Elga and John Hawthorne. 2004. Bayesianism, Infinite Decisions and Binding. *Mind* 113:251-283.

Bostrom, Nick. 2007. Sleeping Beauty and Self Location: A Hybrid Model. *Synthese* 157:59-78.

Carter, Brandon. 1983. The Anthropic Principle and its Implications for Biological Evolution. *Philosophical Transactions of the Royal Society of London. Series A, Mathmatical and Physical Sciences.* 310:347-363.

Colyvan, Mark. 2006. No expectations. *Mind* 115:695-702.

Colyvan, Mark. 2008. Relative Expectation Theory. *Journal of Philosophy* 105(1):37-44.

Easwaran, Kenny. 2008. Strong and weak expectations. *Mind* 117:633-641.

Hájek, Alan. 2006. In Memory of Richard Jeffrey: Some Reminiscences and Some Reflections on *The Logic of Decision. Philosophy of Science* 73:947-958.

Hájek, Alan. 2014. Unexpected Expectations. *Mind* 123:533-567.

Hájek, Alan and Harris Nover. 2006. Perplexing Expectations. *Mind* 115:703-720.

Lewis, Antony. 2001. Comparing cosmological theories. Online. Available at http://cosmologist.info/anthropic.html.

McGee, Vann. 1999. An airtight Dutch Book. *Analysis.* 59:257-265.

Nover, Harris and Alan Hájek. 2004. Vexing Expectations. *Mind* 113:237-249.

Smith, N. 2014. Is evaluative compositionality a requirement of rationality? *Mind* 123:457-502.