FACT SHEET RESEARCH ON BAYESIAN DECISION THEORY

H.R.N. VAN ERP, R.O. LINGER, AND P.H.A.J.M. VAN GELDER

Abstract. In this fact sheet we give some preliminary research results on the Bayesian Decision Theory. This theory has been under construction for the past two years. But what started as an intuitive enough idea, now seems to have the makings of something far more fundamental.

1. Introduction

Thanks to the endeavors of Knuth and Skilling, it has been shown that the product and sum rules of both the Bayesian probability and the Bayesian information theories are derivable by consistency constraints on the lattices of, respectively, statements and questions, [14, 39, 55]; the implication being that in our plausibility and relevancy judgments we humans have a preference for consistency, or, equivalently, rationality. Moreover, Knuth is now researching if the very laws of Nature themselves may be derived by way of consistency constraints on lattices of events, [41].

What we now perceive to be the laws of physics are nothing more than conjectures. These conjectures have obtained the status of laws because of, one the hand, their close correspondence with empirical fact, and, on the the other hand, their power to predict physical phenomena, other than the ones that guided us to these conjectures in the first place. Knuth, and his MaxEnt-colleagues[1], are now in the process of deriving the theorems of Nature, from, what then would be, the primary first principle of Nature itself, that is, consistency.

In light of both these exciting new developments and the fact that these authors, after two years of continuous research, have reached the point that they have come to trust their Bayesian decision theory, to almost the same extent as they have grown to trust the Bayesian probability and information theories[2], these authors have come to entertain the notion that maybe their Bayesian decision theory, which initially started as an intuitive enough Bayesian alternative for the paradigm of behaviorist economics, might actually be Bayesian in the strictest sense of the word; that is, an inescapable consequence of the desideratum of consistency.

[1]MaxEnt-Bayesians are those Bayesians that trace their statistical lineage from Jaynes, back to Jeffreys, back to Laplace.

[2]The former being their field of expertise, and the latter being the subject matter of the first author's current thesis work.

In this fact sheet we will give the eight supporting contacts, in chronological order, that led us to this daring[3] conjecture, together with the lattice theoretical proof of the Bernoulli law, also known as the Weber-Fechner law of sense perception, which until now was the only degree of freedom in our decision theoretical algorithm.

2. The Bayesian Decision Theory

The Bayesian decision theory is very simple in structure. Its algorithmic steps are the following:

(1) Use the product and sum rules of Bayesian probability theory to construct outcome probability distributions.

(2) If our outcomes are monetary in nature, then by way of the Bernoulli law we may map utilities[4] to the monetary outcomes of our outcome probability distributions.

(3) Maximize either the lower or upper bounds, depending on the specific context of the problem of choice we are studying, of the resulting utility probability distributions.

This, then, is the whole of the Bayesian decision theory.

3. Bernoulli's Expected Utility Theory

If we compare the Bayesian decision theory with Bernoulli's initial 1738 expected utility theory, [6]. Then we find that these theories only differ, in that the former proposes to maximize the upper and lower bounds of the utility probability distributions, whereas the latter proposes to maximize their expectation values.

But, expected utility theory, as most economists and economic behaviorists will probably know, is plagued by the Ellsberg and Allais paradoxes, [3, 4, 11]. These paradoxes all boil down to the same two paradoxes.

Problems of choice in which the expected utilities under two hypothetical bets are either the same, implying a lack of preference of one bet over the other, but for which there, nonetheless, is an observable preference for one the bets; or the expected utilities under two hypothetical bets are different, implying a of preference for the bet with the highest expectation value, but where, nonetheless, the bet with the lower expectation value is to be preferred, both empirically and introspectively.

[3]For we share Jaynes' weary and wariness, when he states, [24]: "[W]e have seen enough ambitious but short-lived efforts with the generic title: 'A New Foundation For . . . ' to become a bit weary of them. And we have seen enough putative 'foundations' develop a fluid character unlike real foundations and themselves to the unyielding practical realities, to become a bit wary of them."

[4]Or, as Bernoulli called them, moral values, [6].

For example, in one particular Allais paradox we may be asked which bet we would prefer: The bet in which with certainty we will obtain one million euro, or the bet in which with a probability of 0.5 we will obtain nothing and with a probability of 0.5 we will obtain two million euro. It is found, that we overwhelmingly will opt for the first 'bet'; even though both bets have an expectation value of one million.

If we take the 1-sigma upper and lower bounds, that is,

$$(\mu - \sigma, \mu + \sigma)\,,$$

of both outcome distributions under considerations:

$$\text{CI-bet}_1 : (1.000.000, 1.000.000)\,, \qquad \text{CI-bet}_2 : (0, 2.000.000)\,.$$

Then we see that the first bet maximizes the lower bound, one million relative to zero euro; whereas the second bet maximizes the upper bound, two million relative to one million euro.

So, if our introspection suggests that we ought to take the first bet, then we see that for this particular problem of choice we opt for a lower bound maximization[5]. Moreover, we see that if we take as our criterion of action the maximizations of the confidence bounds, as opposed to the maximizations of the expectation values, that the Allais paradox simply vanishes.

But, if the expectation value maximizations are so problematic, and the solution is so trivial, why did Bernoulli himself not come upon the idea of confidence bound maximization?

Well, the reason that Bernoulli did not, or better yet, could not, take the higher order cumulants of the utility probability distributions into account in his decision theoretical program, was simply because the prerequisite statistical language to think along the lines of confidence bound maximizations was still lacking at the time, when he wrote his 1738 essay[6].

4. First Supporting Contact: case study I.

We now turn to the rationale of the individual to take out insurance and the rationale of the insurance company to provide a single insurance contract. The example given here is a generalization of the insurance example given by Jaynes, (pages 400-402, [28]). It is a generalization in that we now do not compare the

[5]Allais constructed this particular example to demonstrate the psychological reality of variance preferences. People not only try to maximize the expectation value of utility, they also take into account the variances of the respective utility probability distributions. Hence, the name variance preferences, that is, preferences between decisions based upon the variance, or, equivalently, the standard deviations of the utility probability distributions; see Appendix A.

[6]See Appendix B.

means of the utility probability distributions, but, rather, the lower bounds of the utility distributions[7].

In close analogy with Jaynes' treatment of this decision theoretical problem, we provide the reader with a series expansions in the cumulants of the bounds of the premium P. This series expansion allows us to take a look into the inner workings of the black-box, that is the decision theoretical algorithm.

4.1. **The insurance case.** Let P be the premium for some proposed insurance contract between one individual customer and an insurance company. Let C_i for $i = 1, \ldots, n$ enumerate the contingencies covered. The k^{th} contingency has a probability p_k and a cost of L_k if it were to happen. We assume that the insurance company and potential customer make the same probability and costs assessments, p_i and L_i, for the n contingencies C_i.

For both notational simplicity and computational tractability, we will let the probabilities for the contingencies as well as their associated costs be equal, that is,

$$p = p_1 = \cdots = p_n, \qquad L = L_1 = \cdots = L_n.$$

The outcomes can then be defined as

$$O_i = \text{ } i \text{ contingencies occur in conjunction,} \tag{4.1}$$

for $i = 0, 1, \ldots, n$.

For equal probabilities of the respective contingencies the associated probabilities of i contingencies in conjunction follow a Binomial distribution

$$p(\, O_i |\, D_j) = \binom{n}{i} p^i \left(1 - p\right)^{n-i}, \tag{4.2}$$

for decisions $j = 1, 2$.

Let S be the initial wealth, ΔS be the increment in wealth, and ψ be the minimum threshold for monetary stimuli[8]. Then the Bernoulli law for the moral value of the monetary stimulus ΔS is given as:

$$u(\, \Delta S |\, S) = q \log \frac{S + \Delta S}{S}, \qquad -S + \psi < \Delta S < \infty, \tag{4.3}$$

where we note that the Bernoulli law of the moral value of objective monies is the same as the Fechner-Weber law of psycho-physics[9].

[7]Jaynes' insurance example has been instrumental in the formulation of the decision theoretical algorithm, which is given in this fact sheet. As it provided us, at the very start of this research project, almost two years ago, with a blue-print for the treatment of this first case study.

[8]The minimum threshold ψ represents the amount of wealth which constitutes financial ruin.

[9]As an aside, even though psychology traces it roots back to 1789, which is when Kant wrote his metaphysical works on sensing. It was only as late as 1860 that psychology saw its status elevated, from being merely a metaphysical past time, to being a legitimate mathematical and experimental science, [15]. This elevation came about when Fechner applied Bernoulli's utility function as a

4.2. **The premium lower bound for the insurer.** We will now discuss the construction of the utility probability distributions of the insurance company under the decisions to insure and not to insure, respectively, D_1 and D_2. Then we will discuss the rationale lower bound maximizations for this particular instance. After which we will give the intuitive premium lower bound for the insurance company.

Let the insurance company have an initial wealth of M. If the customer pays the insurance premium P and i contingencies occur in conjunction, then the increment in the amount of money for a given outcome O_i is, (4.1),

$$\Delta M_i = P - iL \tag{4.4}$$

Then by way of (4.3), (4.2) and (4.4), we may obtain the following utility probability distribution[10] for the decision to insure D_1,

$$p(u|\, P, D_1) = \sum_{i=0}^{n} \delta\left(u - q \log \frac{M + P - iL}{M}\right) \binom{n}{i} p^i \left(1 - p\right)^{n-i}. \tag{4.5}$$

where δ is the Dirac delta function:

$$\delta(u - c)\, du = \begin{cases} 1, & u = c \\ 0, & u \neq c \end{cases} \tag{4.6}$$

or, equivalently,

$$\int \delta(u - c)\, f(u)\, du = f(c)\,, \tag{4.7}$$

Note that it is property (4.7) of the Dirac-delta function, which enables us to make a one-on-one mapping, from outcomes to utilities.

If the insurance company decides not to sell the insurance, that is, decision D_2, then for each number i of contingencies occurring within the same time period, we have that the initial wealth M of the insurance company remains as is.

So, the corresponding utility probability distribution is[11],

$$p(u|\, D_2) = \delta(u)\,, \tag{4.8}$$

or, equivalently, a probability one of neither loss nor gain.

We now have constructed the utility probability distributions under both decisions D_1, insure, and D_2, do not insure; respectively, (4.5) and (4.8). This leaves us with the choice whether to maximize the lower or upper bounds under the decisions D_1 and D_2.

---

model for the way we humans perceive increments in sensory stimuli; our decibel scale, for example, follows from the Bernoulli law.

[10]See Appendix C.

[11]See Appendix C.

If we maximize the lower bound under the decision D_1, to insure, relative to the decision D_2, not to insure, in the premium P, that is,

$$E(u|\, P, D_1) - \text{std}(u|\, P, D_1) > E(u|\, D_2) - \text{std}(u|\, D_2)\,. \tag{4.9}$$

Then we find the minimum premium P for which the insurer with a high probability[12] will be better of under D_1, relative to decision D_2.

Stated differently, if we solve the premium P under the equality

$$E(u|\, P, D_1) - \text{std}(u|\, P, D_1) = E(u|\, D_2) - \text{std}(u|\, D_2)\,, \tag{4.10}$$

we find the premium P for which the insurance company is indifferent to both the providing and not providing of an insurance.

As the left-hand side of (4.10) becomes much larger than its right-hand side, that is, as the equality (4.10) goes to the inequality, (4.9),

$$E(u|\, P, D_1) - \text{std}(u|\, P, D_1) >> E(u|\, D_2) - \text{std}(u|\, D_2)\,, \tag{4.11}$$

then the insurance company will feel ever more compelled to actively sell his insurance policy to the customer. Since the selling of this policy will tend, as (4.9) tends to (4.11), to a profit net-return in utility, which maps one-to-one to a profit net-return in monetary outcomes. So, for the insurance company the decision inequality (4.9) constitutes an instance of risk-averse, that is, defensive, profit-seeking.

Now, were our insurance company a capital investor, then we may have that the decision inequality involves an investment level, say, I which will be solved to maximize the upper bound of the utility probability distribution under the decision to invest. This upper bound then represents the maximum net-return which has still has a modest probability of occurring[13].

Here we note that confidence intervals represent both an interplay between outcomes and probabilities.

For example, excessive increments in monetary outcomes, having only very small probabilities of occurring, may be weighted by a confidence interval exactly the same as more modest increments in monetary outcomes, which have a much larger likelihood of occurring.

The same also holds for the expectation values, as these values also represent an interplay between outcomes and probabilities, or, equivalently, risk.

Also note that the constant q in (4.5) is the unknown scaling constant of a monetary stimulus. As it turns out, all reference to this constant falls away the

[12]For a normal utility distribution, the region to the right of the lower bound represents a probability of $P(u \geq \mu - \sigma) = 0.84$.

[13]For a normal utility distribution, the region to the right of the upper bound represents a probability of $P(u \geq \mu + \sigma) = 0.16$.

moment we solve inequality (4.9) for the premium P. This holds for all our bound maximizations. This can be seen as follows.

Let X and Y both be stochastic and q some positive constant. Then the inequality

$$E(qX) - \mathrm{std}(qX) > E(qY) - \mathrm{std}(qY) \tag{4.12}$$

is equivalent to the inequality, [43],

$$q\left[E(X) - \mathrm{std}(X)\right] > q\left[E(Y) - \mathrm{std}(Y)\right]. \tag{4.13}$$

Dividing both sides of (4.13) by the constant q, we are left with a further equivalence

$$E(X) - \mathrm{std}(X) > E(Y) - \mathrm{std}(Y), \tag{4.14}$$

in which all mention of the unknown scaling constant q has fallen away.

First we assume that the initial wealth of the insurance company M is much larger than the total damage incurred, should all the contingencies occur in the same time period, that is, nL. Then the logarithm in (4.5) will become linear in the neighborhood of increments of the size nL.

We then compute the mean and standard deviations of the utility probability distributions (4.5) and (4.8), substitute them in (4.9), and solve for the premium P. This gives the following premium lower bound for the insurance company:

$$P > npL + \sqrt{np\,(1-p)}L. \tag{4.15}$$

Seeing that

$$E(iL) = npL \qquad \mathrm{std}(iL) = \sqrt{np\,(1-p)}L, \tag{4.16}$$

we have that (4.15) tells us that the premium must be larger than the 1-sigma monetary damage upper bound, for the insurance company to offer the insurance contract:

$$P > E(iL) + \mathrm{std}(iL). \tag{4.17}$$

This result is pleasantly intuitive, since in the decision theoretical inequality (4.9) a 1-sigma level security level was specified.

Note that, by solving for the premium P, we have gone from the utility dimension, on which our inequality (4.22) was defined, again to the monetary outcome dimension, on which the premium P resides.

4.3. **The premium upper bound for the customer.** There are two distinct decisions for the customer

$$d_1 = \text{Buy Insurance}, \qquad d_2 = \text{Do Not Buy Insurance}.$$

Let the customer have an initial amount of money m. If the customer buys the insurance, D_1, then for any number of contingencies O_i the monetary outcome will always the same. The customer pays the premium P and now has an updated amount of wealth $m - P$, and whatever the number of contingencies, his damages are always refunded to the level $m - P$.

Stated differently, if the customer does buy the insurance, then

$$\Delta m_P = m - P \tag{4.18}$$

Then by way of (4.3), (4.2) and (4.18), we have:

$$p(u|\, d_1) = \delta\left(u - q \log \frac{m - P}{m}\right) \tag{4.19}$$

which is equivalent to the statement that

$$P\left(u = q \log \frac{m - P}{m}\right) = 1.$$

Now, if the customer decides not to buy insurance, d_2, then for a given number of contingencies O_i the monetary damage is iL. So, under O_i the updated amount of wealth is

$$\Delta m_{\bar{i}} = m - iL. \tag{4.20}$$

Then by way of (4.3), (4.2) and (4.20), we have:

$$p(u|\, d_2) = \sum_{i=0}^{n} \delta\left(u - q \log \frac{m - iL}{m}\right) \binom{n}{i} p^i \, (1-p)^{n-i}\,, \tag{4.21}$$

which is equivalent to the statement that

$$P\left(u = q \log \frac{m - iL}{m}\right) = \binom{n}{i} p^i \, (1-p)^{n-i}\,.$$

We now have constructed the utility probability distributions under both decisions d_1, take out an insure, and d_2, do not take out an insurance; respectively, (4.19) and (4.21). This leaves us with the choice whether to maximize the lower or upper bounds under the decisions d_1 and d_2.

The customer does not wish to take out an insurance in order to make a profit. Rather, he wishes to mitigate a potential loss. So, we maximize the lower bound under the decision d_1, to take out an insurance, relative to the decision d_2, not to take out an insure, in the premium P, that is,

$$E(u|\, P, d_1) - \mathrm{std}(u|\, P, d_1) > E(u|\, d_2) - \mathrm{std}(u|\, d_2)\,. \tag{4.22}$$

By doing so, we find the maximum premium P for which the insurer with a high probability will still be better of under d_1, relative to decision d_2. Note that the unknown constant q in (4.19) and (4.21) will fall away in the decision theoretical (4.9); see (4.12) through (4.14).

Solving (4.22) for the premium P, we obtain the premium upper bound for the customer:

$$P < m \left\{1 - \exp\left[E(m - iL|\, d_2) - \mathrm{std}(m - iL|\, d_2)\right]\right\}. \tag{4.23}$$

For a given insurance problem, the inequality (4.23) can simply be evaluated numerically by computing the mean and standard deviation of (4.21) and substituting the corresponding values into (4.23).

Alternatively, if we assume a moderately rich customer for who we may use the log approximation, (4.21),

$$q \log \frac{m - iL}{m} = q \log \left(1 - \frac{iL}{m}\right) \approx q \left[-\frac{iL}{m} - \left(\frac{iL}{m}\right)^2\right], \tag{4.24}$$

then we may make a series expansion of the right hand side of (4.23) in the cumulants of the generating binomial probability distribution[14].

By doing so, we obtain the following approximation of (4.23):

$$P < E(iL) + g \,\mathrm{std}(iL) + O\left(m^{-3/2}\right). \tag{4.25}$$

where g is the factor which quantifies the effect of the curvature, in neighborhood of the possible monetary damages, of the Bernoulli law,

$$g = \sqrt{1 + \frac{E(iL)}{m} + \gamma \frac{\mathrm{std}(iL)}{m} + \frac{1}{4\mathrm{var}(iL)} \frac{\mathrm{var}[\mathrm{var}(iL)]}{m^2}} - \frac{E(iL)}{m}, \tag{4.26}$$

It may be checked, or derived, that as the ratio L/m tends to zero, then g will also tend to one.

Looking at (4.25) and (4.26), we see that the premium upper bound is modulated upward by the spread in monetary damage, the asymmetry in the distribution of monetary damage, and a scaled variance of the variance of the monetary damage; respectively, $\mathrm{std}(iL)$, γ, and $\mathrm{var}[\mathrm{var}(iL)]\,/\mathrm{var}(iL)$.

The latter implies that we are willing to pay a slightly higher premium price if this removes the uncertainty that $\mathrm{var}[\mathrm{var}(iL)]$ entails. Our aversion to loss is not only a function of our uncertainty regarding that loss, but also of the uncertainty we have regarding that uncertainty, [28].

4.4. **Setting a premium on insurances.** In order to set a premium on the insurance contract, we compare the premium constraint of the customer, (4.25), with that of the insurance company, (4.17). It follows that the premium should be in the range

$$E(iL) + \mathrm{std}(iL) < P < E(iL) + g \,\mathrm{std}(iL) \tag{4.27}$$

The margin of profit, MoP, for the insurance company is the premium P minus the range of probable monetary damage $E(iL) + \mathrm{std}(iL)$. Rewriting (4.27), we see that the MoP for the insurance company lies in the range

$$0 < MoP < (g - 1)\,\mathrm{std}(iL) \tag{4.28}$$

[14] This series expansion resulted in some 140 terms, which then had to be condensed manually into the cumulant forms of the binomial probability distribution.

Note that the margin of profit for the insurance company increases as the spread in risk, that is, the spread in monetary damage, $\operatorname{std}(iL)$, increases. This is quite intuitive.

If $\operatorname{std}(iL) = 0$, then both the insurance company will ask as its minimum premium the certain damages, say, kL, where $0 \leq k \leq n$; whereas the customer, on his part, will not be willing to pay more than these certain damages. This then leaves us with a margin of profit of

$$MoP = kL - kL = 0,$$

as predicted by the Bayesian decision theory.

Looking at (4.25) and (4.26), we see that if the maximum amount of monetary damage, nL, is much smaller than the initial amount of wealth m, that is, $nL/m << 1$, then the term g will tend to 1. Consequently, the upper bound of the margin of profit, (4.28), will tend to zero, as the customer becomes his own insurer and there is no profit to be had for the insurance company[15].

Furthermore, the insurance company may exact a considerably larger margin of profit from those that are not moderately rich, that is, for those who have an initial wealth of, say, $m << 1.000.000$. As for them there is a very real chance of suffering financial ruin, were all the contingencies occur at once.

For these cases the approximation (4.25) of (4.23) will break down, as the curvature factor g, (4.26), increases. The premium inequality (4.27) then will go the the corresponding inequality, (4.23):

$$E(iL) + \operatorname{std}(iL) < P < m\left\{1 - \exp\left[E(m - iL|\, d_2) - \operatorname{std}(m - iL|\, d_2)\right]\right\}, \quad (4.29)$$

which expresses a higher willingness by the customer to buy an insurance. As a consequence, (4.28) will go to

$$0 < MoP < m\left\{1 - \exp\left[E(m - iL|\, d_2) - \operatorname{std}(m - iL|\, d_2)\right]\right\} - E(iL) - \operatorname{std}(iL), \quad (4.30)$$

which reflect the subsequent higher margins of profit for the insurance company.

4.5. **Setting a premium on insurances on multiple insurance contracts.** Up to now we have only treated the case of a single contract between an insurance company and just the one solitary customer.

Now, lets assume that the insurance company has N outstanding contracts, each contract covering n contingencies with probability p and a payout of L for each contingency. Then a total of Nn separate contingencies are covered each having probability p. Stated differently, for the insurance company there are now Nn possible outcomes, as opposed to only n possible outcomes. These Nn outcomes

[15]This phenomenon would seem to be a fundamental property of insurances; as it is replicated in our third case study, which also treats an insurance contract of sorts.

have a mean and variance of, respectively,

$$\mu = Nnp = NE(i)\,, \qquad \sigma^2 = Nnp\,(1-p) = N\mathrm{var}(i)\,. \tag{4.31}$$

If we assume that the initial wealth M of the insurance company is sufficient to make its utility for money linear, then it follows that the premium lower bound for the collective of contracts should be set to, (4.15) and (4.31),

$$NP > NnpL + \sqrt{Nnp\,(1-p)}L = NE(iL) + \sqrt{N}\mathrm{std}(iL)\,. \tag{4.32}$$

The lower bound of the insurance company, typically, does not factor into the premium upper bound of the customer. So, this upper bound remains unchanged. It follows that the upper bound of the collective of separate contracts is given by the sum of the N upper bounds:

$$NP < Nm\left\{1 - \exp\left[E(\,m - iL|\,d_2) - \mathrm{std}(\,m - iL|\,d_2)\right]\right\}. \tag{4.33}$$

By way of (4.25), we may for the moderately rich approximate (4.33) as

$$NP < NE(iL) + Ng\,\mathrm{std}(iL)\,, \tag{4.34}$$

where $g \geq 1$, (4.26). Combining (4.32) and (4.34), it follows that the collective margin of profit for the N contracts is, approximately,

$$MoP_N < \sqrt{N}\left(\sqrt{N}g - 1\right)\mathrm{std}(iL)\,. \tag{4.35}$$

For example, for customers with an initial wealth $m = 1.000.000$ of the lower bound of the premium for solitary insurance contracts which cover $n = 10$ contingencies, each having probability of $p = 10^{-4}$ and a maximum payout for each contingency of $L = 50.000$ dollars, is, using (4.16) and the exact[16] (4.23):

$$P > 1631. \tag{4.36}$$

Now, say we have $N = 10.000$ outstanding contracts, then we have that (4.32):

$$NP > 500.000 + 158.100 = 658.100, \tag{4.37}$$

whereas the upper bounds for customer initial wealth $m = 1.000.000$, (4.36), is multiplied by a factor N, giving a collective margins of profit of

$$MoP_N < 16.720.000 - 658.100 = 16.061.900. \tag{4.38}$$

So we see that for a sufficiently large initial wealth M, the law of large numbers combined with the customer's non-linear utility of money allow the insurance company to make a hefty profit on its ten thousand insurance contracts.

[16]Note that the approximation (4.25) will lead to a margin of error of about one and a half dollar, relative to (4.36).

5. Second Supporting Contact: case study II.

We now apply our Bayesian framework to a simple scenario in which a province must decide on how it is willing to invest in a further improvement of its flood defenses. The two decisions under consideration in our case study are

$$D_1 = \text{keep the status quo},$$
$$D_2 = \text{improve the flood defenses}.$$

The investments costs associated with the improvement of the flood defences are designated as

$$I = \text{investment costs associated with improved flood defenses}.$$

The possible outcomes in our risk scenario remain the same under either decision, and as such are not dependent upon the particular decision taken. These outcomes are

$$O_1 = \text{regular river flooding},$$
$$O_2 = \text{catastrophic river flooding},$$
$$O_3 = \text{no flooding},$$

where O_2 is the multiple hazard instance in which the synergy of a regular river flooding in conjunction with a heavy storm conspire to cause a catastrophic flooding.

The decision whether to improve the flood defenses or not is of influence on the probabilities of the respective outcomes. Under the decision to make no additional investments in flood defenses and keep the status quo, D_1, the probabilities of the outcomes will be, say,

$$\begin{aligned} P(O_1|\,D_1) &= 10^{-2}, \\ P(O_2|\,D_1) &= 10^{-5}, \\ P(O_3|\,D_1) &= 1 - P(O_1|\,D_1) - P(O_2|\,D_1)\,. \end{aligned} \tag{5.1}$$

Under the decision to improve the flood defenses, D_2, the probabilities of the flood outcomes will be decreased, leaving us with hypothetical outcome probabilities, say,

$$\begin{aligned} P(O_1|\,D_2) &= 10^{-3}, \\ P(O_2|\,D_2) &= 10^{-7}, \\ P(O_3|\,D_2) &= 1 - P(O_1|\,D_2) - P(O_2|\,D_2)\,. \end{aligned} \tag{5.2}$$

The flood defenses will decrease the chances of a regular river flooding by a factor of only 10. But, as the proposed flood defenses explicitly take into account the failure mechanisms resulting from the simultaneous occurrence of wind storms and a flooding, the chances of a catastrophic river flooding are reduced by a factor 100.

We now proceed to assign utilities to the outcomes. The hypothetical damages associated with the outcomes are, respectively,

$$\begin{aligned} C_1 &= 10 \text{ million euro}, \\ C_2 &= 5 \text{ billion euro}, \\ C_3 &= 0 \text{ euro}. \end{aligned} \tag{5.3}$$

Note that if we were to do an actual analysis, rather than a demonstration of the here proposed decision theoretical framework, then the cost of money itself, in the form of a potential loss of interest on the investment I and the outcomes C_i, should be taken into account also.

Finally, we assume that our province has an initial wealth of M.

Then, by way of the Bernoulli law, (4.3), the utilities for the decision not to invest in additional flood defenses, D_1, are given as:

$$u_i|\, D_1 = q \log \frac{M - C_i}{M}, \qquad i = 1, 2, 3, \tag{5.4}$$

If additional investments are made to improve the flood defenses, D_2, then the utilities become, (4.3):

$$u_i|\, I, D_2 = q \log \frac{M - C_i - I}{M}, \qquad i = 1, 2, 3. \tag{5.5}$$

Seeing that the unknown constant q in (5.4) and (5.5) will fall away in the decision theoretical inequalities, (4.12) through (4.14), we may, without any loss of generality, set $q = 1$.

The utility probability distributions under D_1 and D_2, then can be written out as[17], (5.1), (5.3), (5.9), and (5.4):

$$p(\, u_i|\, D_1) = \begin{cases} P(\, O_1|\, D_1)\,, & u_1 = \log \frac{M - C_1}{M} \\ P(\, O_2|\, D_1)\,, & u_2 = \log \frac{M - C_2}{M} \\ P(\, O_3|\, D_1)\,, & u_3 = \log \frac{M - C_3}{M} = \log \frac{M}{M} = 0 \end{cases}$$

and (5.2), (5.3), (5.9), and (5.5)

$$p(\, u_i|\, I, D_2) = \begin{cases} P(\, O_1|\, D_2)\,, & u_1 = \log \frac{M - C_1 - I}{M} \\ P(\, O_2|\, D_2)\,, & u_2 = \log \frac{M - C_2 - I}{M} \\ P(\, O_3|\, D_2)\,, & u_3 = \log \frac{M - C_3 - I}{M} = \log \frac{M - I}{M} \end{cases}$$

[17]Note that we forgo in this example of the formal Dirac-delta notation. We do this to accentuate the simplicity of the Bayesian decision theoretical algorithm.

where we explicitly condition on the investment I, as this is the variable for which we wish to solve our decision inequalities.

The lower bound of the k-sigma utility confidence interval for the decision to keep the status quo D_1 represents our pre-investment risk exposure. We will only agree to carry the burden of additional investments if this will tend to improve the current risk exposure, or, equivalently, if the lower bound of the k-sigma utility confidence interval under D_2 exceeds the lower bound under D_1.

So, the inequality of interest is the lower bound maximization. By way of the identities, [43],

$$E(X) = \sum_i P(X_i)\ X_i, \qquad \mathrm{std}(X) = \sqrt{\sum_i P(X_i)\ [X_i - E(X)]^2}, \tag{5.6}$$

we may compute the expectations value and standard deviations of (5.6) and (5.6), and construct the lower maximization inequality:

$$E(u|\, I, D_2) - k\ \mathrm{std}(u|\, I, D_2) > E(u|\, D_1) - k\ \mathrm{std}(u|\, D_1)\,, \tag{5.7}$$

We then solve for that investment I where decision D_2 starts to become more profitable than D_1. This I is then the maximal investment we are willing to make in order to improve our flood defenses.

Note that the confidence intervals of (5.6) and (5.6), respectively,

$$[E(u|\, D_1) - k\ \mathrm{std}(u|\, D_1)\,, E(u|\, D_1) + k\ \mathrm{std}(u|\, D_1)]$$

and

$$[E(u|\, I, D_2) - k\ \mathrm{std}(u|\, I, D_2)\,, E(u|\, I, D_2) + k\ \mathrm{std}(u|\, I, D_2)]\,,$$

are used as proxies for the actual utility probability distributions $P(u|\, D_i)$, for $i = 1, 2$; that is, the lower bounds of these confidence bounds in the decision theoretical inequality (5.7) give us a numerical handle on the left-hand position of the utility probability distributions (5.6) and (5.6).

By way of Chebyshev's inequality, [43], we have the following general inequality for the coverage of the k-sigma confidence interval:

$$\text{coverage} = \frac{k^2 - 1}{k^2}. \tag{5.8}$$

If we assume our province to be quite wealthy, with an initial wealth of:

$$M = 10 \text{ billion euro.}$$

Then we find, if we numerically solve for I for different sigma levels k, the following maximal investments for which the implementation of the additional flood defenses is still profitable. These maximal investments, together with the minimal coverages of the corresponding k-sigma intervals, (5.8), are given in Table 1:

sigma level k	coverage CI greater than	maximal investment I
0	n.a.	0.2×10^6
1	0	19.9×10^6
2	3/4	39.5×10^6
3	8/9	59.1×10^6
4	15/16	78.7×10^6
5	24/25	98.1×10^6
6	35/36	117.6×10^6

TABLE 1. maximal I for different sigma levels: $M = 10 \times 10^9$

If we operate on a 6-sigma level of cautiousness, we will be willing to spend up 117.6×10^6 euros for the additional flood defenses which decrease the chances of flooding from (5.1) to (5.2). If we are satisfied with a 1-sigma level of cautiousness, then we are only willing to spend up to 19.9×10^6 euros on those same flood defenses. Note that the maximal investment of 0.2×10^6 euros for $k = 0$ is the expected utility theory solution of this investment optimization problem.

Now, in the previous case study we saw that the Bayesian decision theoretical algorithm has a very intuitive style of reasoning. So, having had our intuition taught in this regard, we will try to reason like the Bayesian algorithm, and then check if we have succeeded in our attempt.

We conjecture that the maximal investment we will be willing to make in the additional flood defenses is the upper bound in the monetary damages under (5.1) and (5.3) minus the upper bound in monetary damages under (5.2) and (5.3).

The former upper bound represents the still probable upper bound in damages if we keep the status quo; whereas the latter upper bound represents the still probable upper bound in damages if we implement additional flood defenses. The distance between these upper bounds represents the margin of investment, where the investments are still expedient.

For a 1-sigma level we find that, (5.3), (5.6), (5.6), and (5.6):

$$E(C_i|\, D_1) + \text{std}(C_i|\, D_1) - [E(C_i|\, D_2) + \text{std}(C_i|\, D_2)] = 14.4 \times 10^6. \tag{5.9}$$

If we assume a linear utility for monetary outcomes, or, equivalently,an initial wealth which is

$$M \to \infty.$$

Then it may be checked that the maximal investment for a 1-sigma level of security goes to:

$$I \to 14.4 \times 10^6,$$

which is as was predicted in (5.9).

Note that again we have that our decision theoretical probability, which was couched in the maximization of the utility of an investment, translates to an investment which is a function of the probabilistic upper bound of the monetary damages.

6. Third Supporting Contact: case study III.

Before premium-based insurances were well and truly introduced in the Northern Netherlands, approximately around the mid-sixteenth century, merchants and ship-owners fell back on different methods for dealing with the financial consequences of long-distance maritime trade. A well known and often applied construction was known as bottomry (bodemerij).

With bottomry a loan was taken out, which was only to be repaid if the vessel or merchandise arrived safely at the port of destination. Therefore, this method incorporated a financing component and was not a pure insurance contract. The premium paid for bottomry (known as opgelt) could amount to as much as 30 or even 70 per cent of the value of the loan, [19].

We will treat the case of bottomry here, using the Bayesian decision theoretic framework. In what follows, we will set the scaling constants q of the Bernoulli law (4.3) to $q = 1$. We may do so, without any loss of generality, since these constants will cancel out in the decision theoretic inequalities, as is demonstrated in (4.12) through (4.14).

6.1. **The insurance case.** We have a merchant with a current wealth of m. The one contingency he wishes to have covered is the loss of his cargo, which would incur a monetary damage of L. If his cargo safely reaches the harbor, the merchant stands to generate a revenue with which he can buy his cargo C times over. Let p be the probability of ship and cargo being lost at sea.

6.2. **The money lender.** The insurer, having an initial wealth of M, will provide the merchant with a bottomry loan of L in exchange for an interest factor c, where $c < C - 1$, to be collectable, together with the loan itself, once the cargo safely reaches the harbor. Under this arrangement, the potential wealth $M_{pot.}$ of the insurer, should the merchant take out a bottomry contract, is

$$M_{pot.} = \begin{cases} M - L, & p \\ M - L + (1 + c)\, L, & 1 - p \end{cases}$$

or, equivalently,

$$M_{pot.} = \begin{cases} M - L, & p \\ M + cL, & 1-p \end{cases} \tag{6.1}$$

By way of (4.3) and (6.1), we may construct the utility probability distribution under E_1, the event that the merchant commits himself to the bottomry contract, as

$$p(u|\, E_1) = \delta\left(u - \log\frac{M-L}{M}\right)p + \delta\left(u - \log\frac{M+cL}{M}\right)(1-p)\,. \tag{6.2}$$

Assuming for the insurer, who is a rich retired merchant, a near linear utility for money in the neighborhood of L, (6.1) simplifies to:

$$p(u|\, E_1) = \delta\left(u + \frac{L}{M}\right)p + \delta\left(u - \frac{cL}{M}\right)(1-p)\,, \tag{6.3}$$

or, equivalently,

$$p(u|\, E_1) = \begin{cases} p, & u = -L/M \\ 1-p, & u = cL/M \end{cases}$$

The insurer will agree with the contract if the lower bound of the utility is greater than zero; zero being the utility of not providing a bottomry loan to the merchant, as no money is lost and no money is gained in that case. This translates to the inequality:

$$E(u|\, E_1) - \mathrm{std}(u|\, E_1) > 0, \tag{6.4}$$

where

$$E(u|\, E_1) = \left[(1-p)\,c - p\right]\frac{L}{M} \tag{6.5}$$

and

$$\mathrm{std}(u|\, E_1) = (1+c)\sqrt{p\,(1-p)}\,\frac{L}{M}. \tag{6.6}$$

Solving inequality (6.4), we find for the insurer the following lower bound for the interest factor c,

$$c > \frac{p + \sqrt{p\,(1-p)}}{(1-p) - \sqrt{p\,(1-p)}}, \tag{6.7}$$

where (6.7), is the 'odds' of the upper bound probability of a ship sinking and the lower bound probability of a ship not sinking; since $\sqrt{p\,(1-p)}$ is the standard deviation of the Bernoulli event of a ship sinking.

One of the central themes of Jaynes' [28] is that Bayesian probability theory, being quantified common sense, may teach our intuition. And it would seem that in (6.7) we have an instance where our intuition is educated by Bayesian decision theory.

These authors already knew that odds are associated with bookmaking[18], so the odds form itself was not that much of a surprise[19]. But the $\sqrt{p\,(1-p)}$ safety correction was not anticipated.

However, this correction, pointed out to us by the Bayesian decision theory, makes nothing but sense, as it expresses the fact that in constructing a conservative odds ratio, that is, a safe bet, one ought to take into account the intrinsic uncertainty we have regarding the occurrence of the Bernoulli event of a ship either sinking or not-sinking.

6.3. **The merchant.** Let m be the amount of money the merchant initially had, before buying his cargo. If the merchant decides to take out the bottomry loan, D_1, then his potential wealth is

$$m_{pot.}|\, D_1 = \begin{cases} m - L + L, & p \\ m - L + L + [C - (1+c)]\, L, & 1-p \end{cases}$$

or, equivalently,

$$m_{pot.}|\, D_1 = \begin{cases} m, & p \\ m + (C - 1 - c)\, L, & 1-p \end{cases} \tag{6.8}$$

where, if the merchant is to get any compensation for all of his hard work, $1 + c < C$, or, equivalently, $c < C - 1$. Under the decision not to hedge against the possible loss of his cargo, D_2, the potential wealth of the merchant is

$$m_{pot.}|\, D_2 = \begin{cases} m - L, & p \\ m - L + CL, & 1-p \end{cases}$$

or, equivalently,

$$m_{pot.}|\, D_2 = \begin{cases} m - L, & p \\ m + (C - 1)\, L, & 1-p \end{cases} \tag{6.9}$$

For the decision to take out the bottomry loan, D_1, the utility distribution may be written down as, (4.3) and (6.8),

$$p(u|\, D_1) = \delta\!\left(u - \log\frac{m}{m}\right) p + \delta\!\left(u - \log\frac{m + (C-1-c)\, L}{m}\right)(1-p)\,. \tag{6.10}$$

[18]As an aside, the pay-out of a winning bet involving regular odds is the money that was put in times a factor $1 + \frac{p}{1-p}$, or, equivalently, $\frac{1}{1-p}$. The insurer, however, demands for his ‘winning bet’ a minimum pay-out of, (6.7): $1 + c = \frac{1}{(1-p)-\sqrt{p(1-p)}}$.

[19]Though this is the first time that we realize that the odds is a decision theoretical measure; as it is a pay-out factor we are to receive if we commit ourselves to a bet. However, a quick literature search on the history of the odds would seem to indicate that odds have traditionally been perceived to be a probability theoretical measure; see also Appendix D.

For the decision not to take out the bottomry loan, D_2, we have, (4.3) and (6.9),

$$p(u|\, D_2) = \delta\left(u - \log\frac{m-L}{m}\right)p + \delta\left(u - \log\frac{m+(C-1)\,L}{m}\right)(1-p)\,. \tag{6.11}$$

The merchant will take out a bottomry loan if the lower bound of the utility under the decision to take out the loan is greater than the lower bound of the utility under the decision not to take out the loan:

$$E(u|\, D_1) - \mathrm{std}(u|\, D_1) > E(u|\, D_2) - \mathrm{std}(u|\, D_2)\,, \tag{6.12}$$

where, (6.10),

$$E(u|\, D_1) = (1-p)\log\frac{m+(C-1-c)\,L}{m}, \tag{6.13}$$

$$\mathrm{std}(u|\, D_1) = \sqrt{p\,(1-p)}\log\frac{m+(C-1-c)\,L}{m}. \tag{6.14}$$

for $c < C-1$; and, (6.11),

$$E(u|\, D_2) = p\log\frac{m-L}{m} + (1-p)\log\frac{m+(C-1)\,L}{m}, \tag{6.15}$$

$$\begin{aligned}\mathrm{std}(u|\, D_2) &= \sqrt{p\,(1-p)}\left|\log\frac{m-L}{m} - \log\frac{m+(C-1)\,L}{m}\right| \\ &= \sqrt{p\,(1-p)}\left(\log\frac{m+(C-1)\,L}{m} - \log\frac{m-L}{m}\right).\end{aligned} \tag{6.16}$$

Solving inequality (6.12), we find the upper bound of the interest factor c, as determined by the merchant:

$$c < \left(C - 1 + \frac{m}{L}\right)\left[1 - \left(\frac{m-L}{m}\right)^{\frac{p+\sqrt{p(1-p)}}{(1-p)-\sqrt{p(1-p)}}}\right]. \tag{6.17}$$

Note that in (6.17) we again encounter the adjusted odds (6.7). Moreover, it may be checked, numerically, that if the merchant himself has an ample fortune, such that $L/m \to 0$, or, equivalently, his utility for money becomes linear, then (6.17) tends to (6.7), that is,

$$\left(C - 1 + \frac{m}{L}\right)\left[1 - \left(\frac{m-L}{m}\right)^{\frac{p+\sqrt{p(1-p)}}{(1-p)-\sqrt{p(1-p)}}}\right] \to \frac{p + \sqrt{p\,(1-p)}}{(1-p) - \sqrt{p\,(1-p)}}. \tag{6.18}$$

So, as the merchant gets richer, the maximum fraction of profit the merchant is willing to share in return for an insurance converges to the minimum fraction of the profit the money lender wishes to receive for that insurance. It follows that the margin of profit for the insurer will evaporate, as the merchants it services become rich enough to become their own insurers. A phenomenon which was also observed in the premium insurance case study.

6.4. **Setting an interest factor on the bottomry loan.** In order to set an interest factor for the bottomry loan, we compare the interest constraint of the customer, (6.17), with that of the insurer, (6.7). We see that the interest factor c that is both acceptable to the insurer and the merchant lies in the range

$$\frac{p+\sqrt{p\,(1-p)}}{(1-p)-\sqrt{p\,(1-p)}} < c < \left(C-1+\frac{m}{L}\right)\left[1-\left(\frac{m-L}{m}\right)^{\frac{p+\sqrt{p(1-p)}}{(1-p)-\sqrt{p(1-p)}}}\right] \tag{6.19}$$

where $c < C-1$, see (6.8).

If the merchant has a cargo which represents a $L = 200$ guilders investment, his total initial wealth being $m = 1.000$ guilders, and a promised return factor of $C = 2$, then we may obtain the following bounds of the interest factor c as a function of the probability of a shipwreck p, Figure 1:

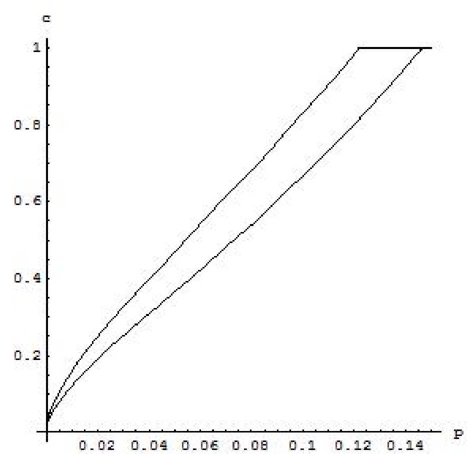

FIGURE 1. Bounds premium factor as function of probability of shipwreck.

From Figure 1, we see that if one in twenty ships gets lost on the high seas, that is, $p = 0.05$, then the minimum interest factor which the insurer will demand in order to cover his risk exposure is $c = 0.37$. The merchant is willing to pay up about half of his net profits, that is, an interest factor of $c = 0.47$, for the bottomry loan of L guilders, as without this loan he stands a small but still very real chance to lose a fifth of his fortune in a shipwreck.

The interest factor c, as a function of the return factor C, is linear in C, (6.7). For $p = 0.05$, $m = 1000$, $L = 200$, we may obtain the following linear equation for the interest factor c:

$$c(C) = 0.314 + 0.078\,C, \tag{6.20}$$

and we see that every unit return factor C increases the merchant willingness with a factor of $\Delta c = 0.078$.

For example, if the return factor is $C = 4$, then our merchant will be willing to pay an interest factor of $c = 0.63$ for the same bottomry contract[20], (6.19), which is an increase in interest c by a factor of 1.33.

So, the increased prospect of his absolute riches make the merchant more inclined to share his wealth, in return for the same commodity; that commodity being a bottomry contract which promises him a riskless profit[21].

In closing, it may observed that as the probability of a maritime mishap approaches $p \to 1/2$, the lower-bound (6.7) will collapse, warning us that inequality (6.4), or, equivalently, risk aversive profit making, tends to become an impossibility. In these cases it might seem that no loans can be had. However, this is not necessarily so.

If the potential profits sufficiently outweigh the potential losses, then risk-seeking venture capitalist may be sought out who are willing to invest in the high-risk, high-yield trade routes.

For instance, the Far East trade at the beginning of the 17th century was both extremely dangerous, with ship loss rates approaching[22] 100%, as well as spectacularly profitable, with initial potential return factors[23] of $C = 50$.

In 1601, following the discovery of the spice sea route, there was a rush on fine spices by Amsterdam merchants. Within the year fourteen expeditions by six different trading companies, sixty-five ships in total, were send around the Cape of Good Hope[24]. But this influx of traders threatened to squeeze the profits right out of the spice trade. In order to remedy the situation the Dutch government established in 1602 a single combined monopoly organization to handle all commerce to the Indies.

---

[20]Note that these interest factors are within the historical interest bounds of 30% to 70%, as reported in [19]. We have tried to distill from [7] and [19] realistic return factors C and probabilities of a ship loss p. Return factors ranging from $C = 2$ to $C = 4$, and a ship-loss frequency of $p = 0.05$, would seem to be reasonable estimates for the 16^{th} century Dutch Levant trade, which was both risky and profitable. But in order to put the decision theoretical model to a more rigorous empirical test, we would have to find a naval historian to collaborate with; seeing that naval history is not our field of expertise.

[21]If so stated, an intuitive enough statement, but it took us a second to arrive at its formulation. We initially computed $c = 0.63$ for $C = 4$, by way of (6.17). This value seemed intuitive enough, more profit means more willingness to pay. So, we did not pay much heed to it. But when we realized that c was linear in C and, consequently, had derived (6.20), we were struck by a sense of initial wonderment.

[22]Prince Henrique the Navigator devoted himself and his families' fortune to finding a sea route around Africa. By the time of Henrique's death in 1460, Portuguese vessels under his patronage had reached the waters of equatorial Africa, but still had not gained the southern passage into the Indian Ocean, [7].

[23]When the *Golden Hind* returned to Plymouth in 1580 laden with the riches of the East, its contents repaid Drake's backers fifty pounds for every one invested, [7].

[24]These expeditions had an aggregated ship loss rate of about 33%.

Investors provided this newly established V.O.C. with 6.5 million guilders in initial funding to hire men, purchase ships, and acquire silver and trade goods to exchange for spices, [7].

So, here we have an historical example of a high-risk and high-yield commercial venture where cheap investment money was easy enough to come by. As both the risks and the profitability of the trade route increases, the moneylender will make the transition from risk-aversive money lending to risk-seeking share holding, and, consequently, the bottomry loan will transition into a capital investment[25].

7. The Skewness Confidence Interval

As stated before, we may compare utility probability distributions by way of their confidence intervals.

By comparing the bounds of the bounds of the confidence intervals of the utility probability distributions, we generalize upon expected utility theory, in that we not only compare the utility expectation values $E(u|\,D_i)$ under the decisions D_i, but also the standard deviations $\text{std}(u|\,D_i)$.

The standard deviations of the utility probability distributions hold pertinent information for our decision problems, as is borne out by the observed phenomena of Source Dependence and Variance Preferences, or, equivalently, the Ellsberg and Allais paradoxes. So, it would stand to reason that the skewness of the utility probability distributions, [23],

$$\text{skew}(u|\,D_i) = \frac{\int \left[u - E(u|\,D_i)\right]^3 p(u|\,D_i)\,du}{\left[\text{std}(u|\,D_i)\right]^3}, \tag{7.1}$$

which is the scaled third order central moment of a distributions, and a measure of its asymmetry, may also hold some pertinent information. And as it turns out, it does. So, we will give here the skewness corrected confidence interval.

If we let, for notational compactness,

$$\mu = E(u|\,D_i)\,, \qquad \sigma = \text{std}(u|\,D_i)\,, \qquad \gamma = \text{skew}(u|\,D_i)\,. \tag{7.2}$$

Then the traditional 1-sigma confidence interval may be written down as:

$$(\mu - \sigma, \mu + \sigma) \tag{7.3}$$

If we let the following three simple considerations be our guide:

(1) The corrected confidence interval should for $\gamma = 0$, this being the skewness of the normal distribution, revert back to (7.3); as it is only by such a property

[25]We conjecture that the Bayesian decision theoretical analyses of capital investments, that is, risk-seeking profit making, will disclose to us a whole new set of dynamics that govern this type of profit making. But though the dynamics might change, what remains constant are the over-arching laws that allow for their expression; those laws being the rules of the Bayesian decision theoretical algorithm.

that the new skewness corrected confidence interval may encompass the standard confidence interval (7.3) as a special limit case.

(2) The corrected confidence interval should take into account the skewness γ in such a way that for $\gamma > 0$ it would compress the lower bound while elongating the upper bound; as this is the qualitative way in which, relative to (7.3), positive skewness ought to be corrected.

(3) The corrected confidence interval should have a coverage for skewed probability distributions that approaches 0.68; as this is the coverage of the sigma confidence interval (7.3) for the non-skewed normal distribution.

Then corresponding skewness corrected 1-sigma confidence intervals are given as, for a skewness of $\gamma > 0$:

$$\left[\mu - \frac{\sigma}{1 + \frac{\sqrt[3]{\gamma}}{1+\gamma+\frac{1}{1+\gamma}}}, \mu + \left(1 + \frac{\sqrt[3]{\gamma}}{1+\gamma+\frac{1}{1+\gamma}}\right)\sigma\right], \tag{7.4}$$

and for a skewness of $\gamma < 0$:

$$\left[\mu - \left(1 - \frac{\sqrt[3]{\gamma}}{1-\gamma+\frac{1}{1-\gamma}}\right)\sigma, \mu + \frac{\sigma}{1 - \frac{\sqrt[3]{\gamma}}{1-\gamma+\frac{1}{1-\gamma}}}\right], \tag{7.5}$$

where it is understood that the third square root of a negative returns a negative.

7.1. **The derivation of the skewness confidence interval.** To the best of our knowledge, no generalization of the time proven interval (7.3), in the form of the intervals (7.4) and (7.5), is to be found in the statistical literature[26]. So, we can sympathize if, at a first glance, these intervals might seem somewhat arbitrary.

[26]Note that Cornish-Fisher expansions are cumulant corrected confidence bounds for sampling statistics, which are obtained by way of series expansions in the sample size n. Here we do not have sampling statistics, samples, or, for that matter, sample sizes; as we ourselves quickly came to realize, in our initial search for the skewness corrected confidence interval. Furthermore, the adjusting terms in the Cornish-Fisher expansions, which are summated, are of the form: power of a cumulant times a polynomial function; which is not the form which we have here, [23].

In order to take away from this possible sense of arbitrariness, we shall now share here the reasoning process that led us to our discovery of the skewness corrected confidence intervals.

The search of (7.4) started with two simple considerations. Firstly, we were looking for a skewness corrected confidence interval which for $\gamma = 0$, this being the skewness of the normal distribution, would revert back to (7.3); as it is only by such a property that the new skewness corrected confidence interval may encompass the standard confidence interval (7.3) as a special limit case. Secondly, we desired from our corrected confidence interval that it should take into account the skewness γ in such a way that for $\gamma > 0$ it should compress the lower bound while elongating the upper bound; as this is the qualitative way in which, relative to (7.3), positive skewness ought to be corrected.

These considerations led us, for positive skewness, to the initial proposal:

$$\left(\mu - \frac{\sigma}{1+\gamma}, \mu + (1+\gamma)\,\sigma\right). \tag{7.6}$$

But it was found that with this proposal the corrected confidence interval of the Bernoulli distributions, for $p \geq 0.5$ and outcomes C_1 and C_2, where $C_1 < C_2$, was approximately constant:

$$\left(\mu - \frac{\sigma}{1+\gamma}, \mu + (1+\gamma)\,\sigma\right) \approx (C_1, C_2)\,, \tag{7.7}$$

with equality holding in the limits $p \to 0.5$ and $p \to 1$.

It was also found that for the binomial distributions, having outcomes $i = 0, 1, \ldots, n$, the interval (7.6) converged to the interval $(0, 1)$, as p, the probability of a success, tended to zero.

This meant that our proposal would not do, for it followed that (7.6) as a confidence interval would lead to a loss of the probabilistic element in our decision theoretical analyses.

Nonetheless, on the up-side, for the exponential distribution,

$$p(x|\,\lambda) = \lambda \exp(-\lambda x)\,, \qquad 0 \leq x < \infty \tag{7.8}$$

which has a mean, standard deviation, and skewness of, respectively,

$$\mu = \frac{1}{\lambda}, \qquad \sigma = \frac{1}{\lambda}, \qquad \gamma = 2, \tag{7.9}$$

there were some encouraging results to report.

The traditional confidence interval of the exponential distribution, as found by way of (7.9) and the unadjusted (7.3), is $\left(0, \frac{2}{\lambda}\right)$ and has a coverage of:

$$\int_0^{\frac{2}{\lambda}} \lambda \exp(-\lambda x)\, dx = 0.86. \tag{7.10}$$

It was found that (7.6), by way of (7.9), translated to the interval $\left(\frac{2}{3\lambda}, \frac{4}{\lambda}\right)$. This interval had a coverage of:

$$\int_{\frac{2}{3\lambda}}^{\frac{4}{\lambda}} \lambda \exp\left(-\lambda x\right) dx = 0.50. \tag{7.11}$$

Now, if we compare the coverages (7.11) and (7.10) with the coverage of the standard sigma interval $(\mu - \sigma, \mu + \sigma)$ for the normal distribution:

$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right] dx = 0.68, \tag{7.12}$$

then it would seem that the adjusted (7.11) was no worse than the traditional (7.10). Moreover, the lower bound of the adjusted interval was no longer the trivial zero, while the upper bound had been elongated by, what would seem to be a reasonable factor.

These modest successes for the confidence interval of the continuous exponential distribution, in terms of qualitative behavior and actual coverage, managed to give us a sense of being on the right track somehow.

We then contemplated that the standard deviation σ is the square root of the second order central moment; the square root being the operation by which we translate the second-order information about the spread in our probability distribution to the first-order dimension, which is the dimension in which our propositions of interest reside.

So maybe we had to take the third central moment $m^{(3)}$, (7.1) and (7.2):

$$m^{(3)} = \int \left(x-\mu\right)^3 p(\,x|\,\{\theta\})\,dx = \gamma\,\sigma^3, \tag{7.13}$$

take its third square root, and then replace the γ's in (7.6) with that root.

This led us to our second proposal

$$\left(\mu - \left(1 + \sqrt[3]{\gamma}\,\sigma\right)\sigma, \mu + \frac{\sigma}{1 + \sqrt[3]{\gamma}\,\sigma}\right), \tag{7.14}$$

But it was found that with this second proposal the corrected confidence interval of the Bernoulli distributions, for $p \geq 0.5$ and outcomes C_1 and C_2, where $C_1 < C_2$, resulted in an unwanted factor $C_2 - C_1$ in the σ's following $\sqrt[3]{\gamma}$.

Without this factor (7.14) seemed to work quite well for the Bernoulli distributions, with corrected confidence intervals that were probabilistic; that is, intervals whose bounds converged to the expectation values as we approached certainty. So, the question then became: How to loose this factor $C_2 - C_1$ in a non-arbitrary manner?

If we could express the factor $C_2 - C_1$ as a function of the cumulants of the Bernoulli distribution, then we could, on the one hand, divide this disruptive factor out and, on the other hand, obtain the, apparently, necessary cumulant correction for our skewness confidence interval.

We then remembered that our initial proposal (7.6), when applied to Bernoulli distributions, resulted in the non-probabilistic interval (C_1, C_2), which has a range of $C_2 - C_1$. This range being equal the factor that we wished to see eliminated.

Rewriting the interval (7.6) as a range, we arrived at the 'support':

$$\mu + (1+\gamma)\,\sigma - \left(\mu - \frac{\sigma}{1+\gamma}\right) = (1+\gamma)\,\sigma + \frac{\sigma}{1+\gamma}. \tag{7.15}$$

Substituting (7.15) into (7.14), in such a way that the factor $C_2 - C_1$ was lost, we then obtained our final proposal (7.4):

$$\left[\mu - \frac{\sigma}{1 + \frac{\sqrt[3]{\gamma}}{1+\gamma+\frac{1}{1+\gamma}}}, \mu + \left(1 + \frac{\sqrt[3]{\gamma}}{1+\gamma+\frac{1}{1+\gamma}}\right)\sigma\right].$$

Having found (7.4), it was then easy enough to find, by way of symmetry arguments, the skewness corrected interval (7.5) for $\gamma < 0$.

7.2. **Supporting contacts for the skewness confidence interval.** The interval (7.4), together with (7.9), translates for the exponential distribution, (7.8),

$$p(x|\,\lambda) = \lambda \exp\left(-\lambda x\right), \qquad 0 \le x < \infty$$

which has a mean, standard deviation, and skewness of, respectively, (7.9),

$$\mu = \frac{1}{\lambda}, \qquad \sigma = \frac{1}{\lambda}, \qquad \gamma = 2,$$

to the skewness corrected interval

$$\left(\frac{3\sqrt[3]{2}}{\left(10 + 3\sqrt[3]{2}\right)\lambda}, \frac{20 + 3\sqrt[3]{2}}{10\lambda}\right). \tag{7.16}$$

This interval has a coverage of:

$$\int_{\frac{3\sqrt[3]{2}}{10+3\sqrt[3]{2}}\frac{1}{\lambda}}^{\frac{20+\sqrt[3]{2}}{10}\frac{1}{\lambda}} \lambda \exp\left(-\lambda x\right) dx = 0.67. \tag{7.17}$$

which is very close to the benchmark coverage value of 0.68 of the 1-sigma confidence interval of the normal distribution, (7.12),

$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right] dx = 0.68.$$

We conjecture that the missing 0.01 probability density coverage in (7.17), relative to (7.12), is a function of the kurtosis and the other higher order cumulants of the exponential distribution[27]; seeing that any probability distribution is wholly determined by its moments.

In contrast, the traditional confidence interval of the exponential distribution, as found by way of (7.9) and the unadjusted (7.3), is

$$\left(0, \frac{2}{\lambda}\right) \tag{7.18}$$

and has a coverage of, (7.10),

$$\int_0^{\frac{2}{\lambda}} \lambda \exp\left(-\lambda x\right) dx = 0.86.$$

So, it would seem that the adjusted (7.16) is much closer to the mark than the traditional (7.18).

The beta distribution is defined as:

$$p(\theta|\, r, n) = \frac{(n-1)!}{(r-1)!\,(n-r-1)!}\theta^{r-1}\left(1-\theta\right)^{n-r-1}, \qquad 0 \leq \theta \leq 1 \tag{7.19}$$

For $r = 5$ and $n = 10$, where we have a symmetrical beta distribution with $\gamma = 0$, and an expectation value of

$$E(\theta) = \mu = \frac{r}{n} = 0.5, \tag{7.20}$$

we have that (7.4) collapses, by construction, to (7.3), giving a shared confidence interval of

$$\left(0.35, 0.65\right),$$

which corresponds with a coverage of

$$\int_{0.35}^{0.65} \frac{(n-1)!}{(r-1)!\,(n-r-1)!}\theta^{r-1}\left(1-\theta\right)^{n-r-1} d\theta = 0.66. \tag{7.21}$$

A coverage which, for $r = n/2$, will converge to the benchmark coverage (7.12), as n goes to infinity.

For the more severe case of $r = 1$ and $n = 10$, where we have a skewed beta distribution, $\gamma = 1.47$, and an expectation value

$$E(\theta) = \mu = \frac{r}{n} = 0.1,$$

we find that (7.4) will give the corrected confidence interval

$$\left(0.04, 0.23\right),$$

[27] We also conjecture that the kurtosis correcting term will be exponentially larger than the skewness correcting term; seeing the progression from the standard deviation 'correcting' term, to the skewness correcting term:

$$\sigma \qquad \text{to} \qquad 1 + \frac{\sqrt[3]{|\gamma|}}{1 + |\gamma| + \frac{1}{1+|\gamma|}}.$$

which corresponds with a coverage of

$$\int_{0.04}^{0.23} \frac{(n-1)!}{(r-1)!\,(n-r-1)!} \theta^{r-1} (1-\theta)^{n-r-1} \, d\theta = 0.63, \tag{7.22}$$

which is still very close to the benchmark (7.12).

For comparison, for $r = 1$ and $n = 10$ the uncorrected (7.3) will give a confidence interval,

$$(0.01, 0.19),$$

whose lower bound is four times closer to the trivial zero, and as a consequence, as the bulk of the probability density of a positively skewed distribution lies to the left, will give an inflated coverage of

$$\int_{0.01}^{0.19} \frac{(n-1)!}{(r-1)!\,(n-r-1)!} \theta^{r-1} (1-\theta)^{n-r-1} \, d\theta = 0.77. \tag{7.23}$$

Furthermore, for $r = 1$ and $n \to \infty$, the skewness of the beta distribution converges to

$$\gamma \to 2.$$

As a consequence, the coverage of unadjusted interval (7.3) diverges from the benchmark coverage (7.12), with a 'limit' of 0.86, for $n = 10^6$. In contrast, the coverage of (7.4) converges to the coverage (7.17), with a 'limit' of 0.67, for $n = 10^6$; where we note that the exponential distribution (7.8) has a skewness of $\gamma = 2.$, (7.9), and a convergence, for the sigma interval of (7.10).

So, we find that the skewness corrected confidence intervals, for both exponential and beta distributions, give us excellent coverages which are extremely close to the benchmark coverage of the normal distribution, or tend to do so, in some well-defined limit.

8. Fourth Supporting Contact: A Severe Test

With the skewness adjusted intervals (7.4) and (7.5), we have extended the scope of our theory, in that we now not only take into account the means and standard deviations of the utility probability distributions, but also their skewnesses.

As the distributions in our previous case studies were all highly skewed, we will take a quick look at the practical implications that the skewness adjusted intervals hold for the worked out bottomry contract example.

We do this by giving some summary results of the re-analyzed example where, instead of using (7.3),

$$(\mu - \sigma, \mu + \sigma), \tag{8.1}$$

we will now use, depending the sign of the skewness γ, either (7.4),

$$\left[\mu - \frac{\sigma}{1 + \frac{\sqrt[3]{\gamma}}{1+\gamma+\frac{1}{1+\gamma}}}, \mu + \left(1 + \frac{\sqrt[3]{\gamma}}{1+\gamma+\frac{1}{1+\gamma}}\right)\sigma\right], \qquad \gamma > 0, \tag{8.2}$$

or, (7.5),

$$\left[\mu - \left(1 - \frac{\sqrt[3]{\gamma}}{1-\gamma+\frac{1}{1-\gamma}}\right)\sigma, \mu + \frac{\sigma}{1 - \frac{\sqrt[3]{\gamma}}{1-\gamma+\frac{1}{1-\gamma}}}\right], \qquad \gamma < 0. \tag{8.3}$$

8.1. **A re-analysis of case study III.** Let p be the probability of ship and cargo being lost at sea. With a bottomry contract, a loan was taken out, which was only to be repaid if the vessel or merchandise arrived safely at the port of destination. The premium paid for bottomry could amount to as much as 30 or even 70 per cent of the value of the loan, [19].

On the one hand we have a merchant with a current wealth of m. The one contingency this merchant wishes to have covered is the loss of his cargo, which would incur a monetary damage of L. If his cargo safely reaches the harbor, the merchant stands to generate a revenue with which he can buy his cargo C times over.

On the other hand we have an insurer, having an initial wealth of M, who will provide the merchant with a bottomry loan of L in exchange for an interest factor c, where $c < C - 1$, to be collectable, together with the loan itself, once the cargo safely reaches the harbor.

Because of the way we have set up our problem, that is, labeled our outcome labels, both the insurer and the merchant have negative skewnesses. So, we have to use (8.3) for the construction of the skewness corrected lower confidence bounds of the utility probability distributions.

If, for the insurer, we substitute the skewness corrected lower bounds of the utility probability distributions of the decisions E_1, provide a bottomry loan, and E_2, do not provide a bottomry loan, into the decision inequality (6.4), and solve for the interest factor c, we obtain as the lower bound of the interest factor, as determined by the insurer:

$$c > \frac{p + \left(1 + \frac{\sqrt[3]{\frac{1-2p}{\sqrt{p(1-p)}}}}{1+\frac{1-2p}{\sqrt{p(1-p)}} + \frac{1}{1+\frac{1-2p}{\sqrt{p(1-p)}}}}\right)\sqrt{p(1-p)}}{(1-p) - \left(1 + \frac{\sqrt[3]{\frac{1-2p}{\sqrt{p(1-p)}}}}{1+\frac{1-2p}{\sqrt{p(1-p)}} + \frac{1}{1+\frac{1-2p}{\sqrt{p(1-p)}}}}\right)\sqrt{p(1-p)}}, \tag{8.4}$$

where we, for now, have propagated the minus sign in the skewness of the utility probability distribution under decision E_1, that is,

$$\gamma = -\frac{1-2p}{\sqrt{p(1-p)}}, \tag{8.5}$$

through the lower bound of (8.3).

In the case of a Bernoulli probability distribution having an event probability p and outcomes 0 and 1, we have that the first three cumulants are given as:

$$\mu = p, \qquad \sigma = \sqrt{p(1-p)}, \qquad \gamma = \frac{1-2p}{\sqrt{p(1-p)}}. \tag{8.6}$$

Since we have that for this Bernoulli probability distribution the skewness is positive for $p < 0.5$, we may readily check that the numerator in (8.4) corresponds with the skewness corrected upper bound probability of a Bernoulli probability distribution having probability p, (8.2) and (8.6).

In the case of a Bernoulli probability distribution having an event probability $1-p$ and outcomes 0 and 1, we have that the first three cumulants are given as:

$$\mu = 1-p, \qquad \sigma = \sqrt{p(1-p)}, \qquad \gamma = -\frac{1-2p}{\sqrt{p(1-p)}}. \tag{8.7}$$

As for this Bernoulli probability distribution the skewness is negative for $p < 0.5$, we will, for demonstrative purposes, allocate the minus signs again to their corresponding skewness, and so restore the initial minuses.

If we do so, we get

$$c > \frac{p + \left(1 + \frac{\sqrt[3]{\frac{1-2p}{\sqrt{p(1-p)}}}}{1+\frac{1-2p}{\sqrt{p(1-p)}} + \frac{1}{1+\frac{1-2p}{\sqrt{p(1-p)}}}}\right)\sqrt{p(1-p)}}{(1-p) - \left(1 - \frac{\sqrt[3]{-\frac{1-2p}{\sqrt{p(1-p)}}}}{1-\left(-\frac{1-2p}{\sqrt{p(1-p)}}\right) + \frac{1}{1-\left(-\frac{1-2p}{\sqrt{p(1-p)}}\right)}}\right)\sqrt{p(1-p)}}. \tag{8.8}$$

from which it may now be readily checked, as (8.8) and (8.4) are equivalent, differing only in a retraction of the minus signs, that the the denominator in (8.4) corresponded with the skewness corrected lower bound probability of of a Bernoulli probability distribution having probability $1 - p$, (8.3) and (8.7).

It follows that our initial interpretation of the lower bound of the interest factor c being an adjusted odds for a Bernoulli event occurring, remains in (8.4), or, equivalently, (8.8), just as valid as it was in (6.7); though, with all the nested skewnesses, we may loose the immediacy of this recognition.

Moreover, we see that the Bayesian decision theory automatically does the right thing, two steps ahead of own faltering intuition. As we ourselves experienced, when first constructing the skewness adjusted odds.

Initially we subtracted the term $(1 - p)$ in (6.7) minus the skewness corrected standard deviation of the lower bound of (8.2), which belongs to p. But the Bayesian decision theory then reminded us, by collapsing the model, that this was misguided.

In our initial problem formulation, in terms of utilities, the skewness was negative, leading to the adjusted lower bound as given in (8.3).

The Bayesian decision algorithm, then, in the finding of (8.8), reformulates the problem as the computing of an adjusted odds ratio, where two distinct Bernoulli probability distributions deliver us the probabilities of the events of a ship sinking and a ship not-sinking.

The Bernoulli event of a ship sinking, for $p < 0.5$, has a positive skewness, (8.6), which gives an upper bound adjustment, (8.2), which is equivalent to the lower bound adjustment of the initial problem formulation, which had a negative skewness, (8.3).

Whereas, the Bernoulli event of a ship not-sinking, for $p < 0.5$, has a negative skewness, (8.7), which gives an lower bound adjustment, (8.3), which is equivalent to the lower bound adjustment of the initial problem formulation, which had a negative skewness, (8.3).

For the merchant we substitute the skewness corrected lower bounds of the utility probability distributions of the decisions D_1, take out a bottomry loan, and D_2, do not take out a bottomry loan, into the decision inequality (6.12). If we then solve for the interest factor c, we obtain the upper bound of the interest factor, as determined by the merchant:

$$c < \left(C - 1 + \frac{m}{L}\right)\left[1 - \left(\frac{m - L}{m}\right)^{f(p)}\right], \tag{8.9}$$

where

$$f(p) = \frac{p + \left(1 + \frac{\sqrt[3]{\frac{1-2p}{\sqrt{p(1-p)}}}}{1+\frac{1-2p}{\sqrt{p(1-p)}} + \frac{1}{1+\frac{1-2p}{\sqrt{p(1-p)}}}}\right)\sqrt{p\,(1-p)}}{(1-p) - \left(1 - \frac{\sqrt[3]{-\frac{1-2p}{\sqrt{p(1-p)}}}}{1-\left(-\frac{1-2p}{\sqrt{p(1-p)}}\right) + \frac{1}{1-\left(-\frac{1-2p}{\sqrt{p(1-p)}}\right)}}\right)\sqrt{p\,(1-p)}}.$$

In Figure 2, we give for the bounds of the interest factor c as a function of the probability of a shipwreck p. This is done for the case where the merchant has a cargo which represents a $L = 200$ guilders investment, a total initial wealth of $m = 1.000$ guilders, and an expected return factor of $C = 2$.

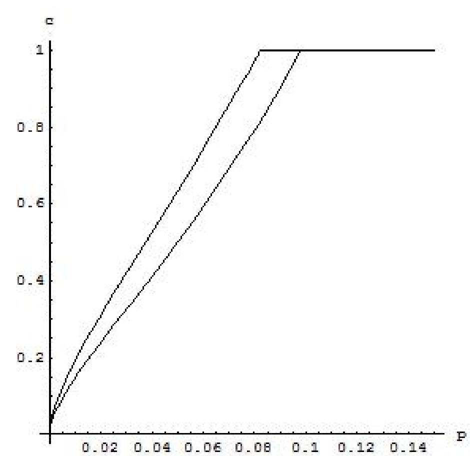

FIGURE 2. Skewness corrected bounds premium factor as function of probability of shipwreck.

From Figure 2, we see that if one in twenty ships gets lost on the high seas, that is, $p = 0.05$, then the minimum interest factor which the insurer will demand in order to cover his risk exposure is $c = 0.50$. So, for $p = 0.05$, the skewness correction has led to an increase of the lower bound of the interest factor c with a factor of 1.35.

The merchant is now willing to pay an interest factor of $c = 0.63$. So, for $p = 0.05$, the skewness correction has led to an increase of the upper bound of the interest factor c with a factor of 1.34.

So, for the insurer more money is to be made, as the merchant now takes the skewness of the Bernoulli distribution into account and, consequently, is willing to pay more interest on his bottomry loan. But at the same time the insurer is also more cognizant of the risk he is taking by providing a bottomry loan, as he himself now also takes the skewness of the Bernoulli distribution into account.

Furthermore, for $p = 0.05$, $m = 1000$, $L = 200$, we may obtain the following skewness corrected linear equation for the interest factor c:

$$c(C) = 0.423 + 0.106\, C, \tag{8.10}$$

and we see that every unit return factor C increases the merchant willingness with a factor of $\Delta c = 0.106$.

The linear equation (8.10) constitutes, relative to the uncorrected case (6.20), a difference of 0.109 on the intercept and a difference of 0.027 on the slope.

So, we again see that the merchant has become more aware of the risks involved, in terms of the possibility of losing a fifth of his fortune, as his willingness to pay a higher interest factor has generally increased.

9. Fifth Supporting Contact: The Psychological Certainty Effect, Part I

Risk seeking refers to a specific pattern in betting behavior. Uncertain larger gains are preferred over sure smaller gains and uncertain larger losses are preferred over sure smaller losses. The psychologists Kahneman and Tversky state that risk seeking constitutes one of the minimal challenges that must be met by any adequate descriptive theory of choice, [60].

The observation that large gains are preferred over sure much smaller gains is commensurate with the fact that we may prefer high-risk, high-yield investment opportunities over low-risk, low-yield ones. Likewise, the observation that uncertain larger losses are preferred over sure smaller, though still substantial, losses is in accordance with those instances in the past where traders incurred hundreds of millions in losses, in their attempts to make good on their previous losses[28].

If the signs of the outcomes in the risk seeking betting scenarios are reversed, then the preferences between the bets will also reverse. This is called the reflection effect, [33]. So, risk seeking in the positive domain is accompanied by risk aversion in the negative domain. Conversely, risk seeking in the negative domain is accompanied by risk aversion in the positive domain.

[28] As, for example, happened to Nicholas William Leeson, a trader for the Barrings Bank in the nineties. Though we believe that Leeson would have acted less recklessly had he been investing his own money, instead that of the deposit holders. That is, we expect that his Weber constant for his own money, say, q, was markedly larger than his Weber constant for the deposit holders money, say, q_0, where $0 \leq q_0 << q$.

We will see that risk seeking corresponds with a predominant tendency to maximize the upper bounds of our utility probability distributions, whereas risk aversion corresponds with a predominant tendency to maximize the lower bounds[29].

9.1. **Risk Seeking I.** We first give an example of risk seeking in the case of a small probability of winning a large prize, that is, risk seeking in the positive domain. This case of risk seeking represents our tendency to profit maximization and demonstrates that we will be willing to invest in a long shot if the pay-out is high enough.

The outcome probability distributions for the respective bets in our risk seeking example are

$$p(\,O|\,D_1) = \begin{cases} 0.001, & O = 5000 \\ 0.999, & O = 0 \end{cases} \tag{9.1}$$

and

$$p(\,O|\,D_2) = \begin{cases} 1.0, & O = 5 \end{cases} \tag{9.2}$$

It is found that 72% of $N = 72$ subjects prefer decision D_1 over D_2, [33]. Even though both bets have the same expectation value of

$$E\left(\,O|\,D_1\right) = 0.001 \times 5000 = 5 = 1.0 \times 5 = E\left(\,O|\,D_2\right).$$

We now interpret this finding in terms of the Bayesian decision theoretic framework.

Kahneman and Tversky state that the median net monthly income for a family is about 3000 Israeli pounds, [33], being that the subjects were all students we will assume an initial amount of money of $m = 1000$ Israeli pounds. This gives us the following utility probability distributions:

$$p(\,u|\,D_1) = \begin{cases} 0.001, & u = q \log \frac{6000}{1000} \\ 0.999, & u = 0 \end{cases} \tag{9.3}$$

and

$$p(\,u|\,D_2) = \begin{cases} 1.0, & u = q \log \frac{1005}{1000} \end{cases} \tag{9.4}$$

Using the identities, [43]:

$$E(X) = \sum_i X_i P_i, \qquad \text{std}(X) = \sqrt{\sum_i \left[X_i - E(X)\right]^2 P_i}, \tag{9.5}$$

and [23]:

$$\text{skew}(X) = \frac{\sqrt{\sum_i \left[X_i - E(X)\right]^3 P_i}}{\left[\text{std}(X)\right]^3}, \tag{9.6}$$

[29]Note that risk aversion is the mechanism which provides the rationale for both the investing in flood defenses and the taking out of an insurance. To be more precise, in the investment example the operating mechanism is risk aversion in the negative domain; whereas in the insurance examples, we have customers who operate on the basis of risk aversion in the negative domain, and insurers who operate on the basis of risk aversion in the positive domain (e.g., defensive profit making).

and, depending on the sign of (9.6), either the skewness confidence interval (7.4) or (7.5), we may construct the skewness corrected intervals:

$$[LB(\,u|\,p_1, D_1)\,, UB(\,u|\,p_1, D_1)]\,, \tag{9.7}$$

and

$$[LB(\,u|\,p_2, D_2)\,, UB(\,u|\,p_2, D_2)]\,. \tag{9.8}$$

Letting D_2 be the decision to choose for the certain gain or loss, that is, $p_2 = 1$, and relabeling $p_1 = p$, we may construct the general decision theoretical equality, (9.7) and (9.8):

$$[LB(\,u|\,1, D_2) - LB(\,u|\,p, D_1) = UB(\,u|\,p, D_1) - UB(\,u|\,1, D_2)]\,. \tag{9.9}$$

If we solve this equality for p, then we find the probability p of the uncertainty bet D_1, for which the bets D_1 and D_2 are undecided; that is, that probability p for which D_1 and D_2 are in fair, in that a larger lower loss bound under D_1 is compensated with a commensurate larger upper gain bound.

If we solve for the probability p, we find the fair probability under bet D_1:

$$p = 0.0000288$$

So, if the probability of the uncertain events exceeds the lower bound

$$p > 0.0000288,$$

then we will accept the uncertain bet D_1. As the gain in the utility upper bound under D_1 will dominate the loss in the utility lower bound under D_2.

It is found that 72% of $N = 72$ subjects prefer decision D_1 over D_2, [33], even though both bets have the same outcome expectation values. The phenomenon of utility upper bound dominance for gains constitutes risk seeking in the positive domain.

We may plot the fair probability p for a certainty bet as a function of the initial wealth m, where we let $200 < m < 10.000$, Figure 3:

As the initial wealth $m \to \infty$, the utility of an increment in wealth in the range of $5 < \Delta m < 5000$ will both become linear and tend to zero, and the fair probability will converge to

$$p \to 0.0000039. \tag{9.10}$$

Note that if we commit ourselves to a value of the constant q, this constant being the appropriate utility scaling factor for monetary outcomes[30], we may construct attraction maps, with increments, both positive and negative, of, say, ten utilities, relative to the probability fairness baseline in Figure 3.

[30]One may obtain a value for q by either personal introspection, or by psychological experimentation, where subjects are asked to report their introspection.

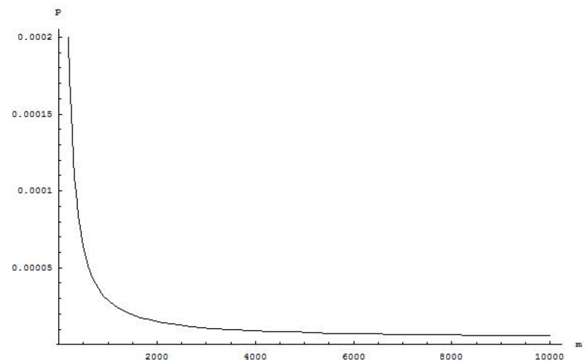

FIGURE 3. Fair probability as function of initial wealth.

9.2. **Risk Aversion I.** The above analysis may also be performed for the case when there is a small probability of loosing a large sum of money. We then will see a reversal in the preference for bet D_1 over bet D_2 to a preference for bet D_2 over bet D_1. Risk aversion in the negative domain represents our tendency to hedge against large and catastrophic losses.

The outcome probability distributions for the respective bets are[31]:

$$p(O|\, D_1) = \begin{cases} 0.001, & O = -5000 \\ 0.999, & O = 0 \end{cases} \tag{9.11}$$

and

$$p(O|\, D_2) = \begin{cases} 1.0, & O = -5 \end{cases} \tag{9.12}$$

It is found that 83% of $N = 72$ subjects preferred the bet D_2 over D_1, [33].

We now will imagine that the students of the Kahneman and Tversky experiments, who were asked to perform imaginary bets, have an imaginary initial amount of money of $m = 6000$ Israeli pounds[32]. Assuming the utility function (4.3), we get the following utility probability distributions:

$$p(u|\, D_1) = \begin{cases} 0.001, & u = q \log \frac{1000}{6000} \\ 0.999, & u = 0 \end{cases} \tag{9.13}$$

and

$$p(u|\, D_2) = \begin{cases} 1.0, & u = q \log \frac{5995}{6000} \end{cases} \tag{9.14}$$

If we solve (9.9) for p, then we find the probability p of the uncertainty bet D_1, for which the bets D_1 and D_2 are undecided:

$$p = 0.0000008.$$

[31]Compare with (9.1) and (9.2).

[32]Kahneman and Tversky do not take the initial wealth m into account in their discussion of their experimental results; see Appendix E.

So, if the probability of the uncertain events exceeds the lower bound

$$p > 0.0000008,$$

then we will accept the certain bet D_2. As the gain in the utility lower bound under D_2 will dominate the loss in the utility upper bound under D_1.

It is found that 83% of $N = 72$ subjects prefer decision D_1 over D_2, [33], even though both bets have the same outcome expectation values. The phenomenon of utility lower bound dominance for losses constitutes risk aversion in the negative domain.

We may plot the fair probability p for a certainty bet as a function of the initial wealth m, where we let $5200 < m < 10.000$, Figure 4:

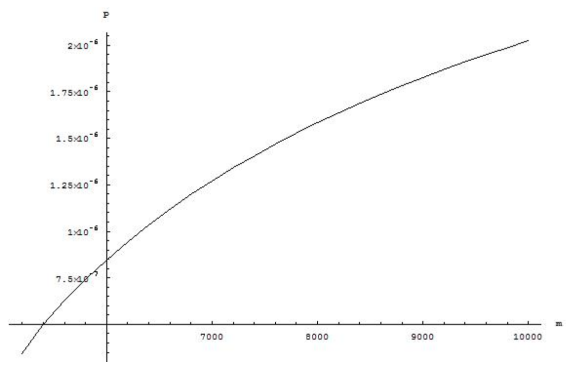

FIGURE 4. Fair probability as function of initial wealth.

As the initial wealth $m \to \infty$, the utility of an increment in wealth in the range of $-5000 < \Delta m < -5$ will both become linear and tend to zero, and the fair probability will converge to the convergance of the symmetrical case, where the outcomes are positive, (9.10),

$$p \to 0.0000039.$$

9.3. **Risk Seeking II.** We now give an example of risk seeking when people must choose between a sure loss and a substantial probability of a larger loss, that is, risk seeking in the negative domain. This case of risk seeking represents our tendency to try to evade large and catastrophic losses.

The outcome probability distributions for the respective bets in our risk seeking example are

$$p(O|\, D_1) = \begin{cases} 0.5, & O = -1000 \\ 0.5, & O = 0 \end{cases} \tag{9.15}$$

and

$$p(O|\, D_2) = \begin{cases} 1.0, & O = -500 \end{cases} \tag{9.16}$$

It is found that 69% of $N = 68$ subjects preferred the bet D_2 over D_1, [33].

We now imagine an initial amount of money of $m = 1500$ Israeli pounds. Assuming the utility function (4.3), we get the utility probability distributions:

$$p(\,u|\,D_1) = \begin{cases} 0.5, & u = q \log \frac{500}{1500} \\ 0.5, & u = 0 \end{cases} \tag{9.17}$$

and

$$p(\,u|\,D_2) = \begin{cases} 1.0, & u = q \log \frac{1000}{1500} \end{cases} \tag{9.18}$$

If we solve (9.9) for p, then we find the probability p of the uncertainty bet D_1, for which the bets D_1 and D_2 are undecided:

$$p = 0.191.$$

We may plot the fair probability p for a certainty bet as a function of the initial wealth m, where we let $1500 < m < 10.000$, Figure 5:

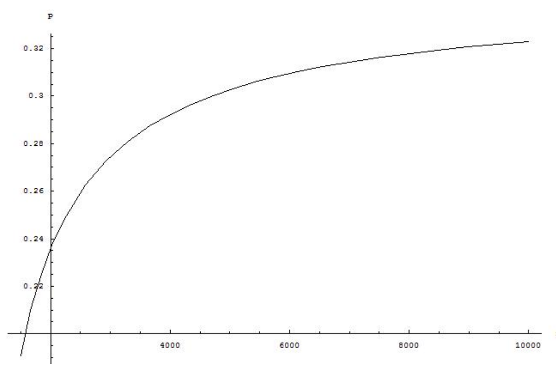

FIGURE 5. Fair probability as function of initial wealth.

As the initial wealth $m \to \infty$, and Bernoulli law converges to

$$q \log \left(\frac{m - O}{m} \right) \to q \frac{O}{m}, \tag{9.19}$$

the utility of an increment in wealth in the range of $-1000 < \Delta m < -500$ will both become linear and tend to zero, and the fair probability will converge, because of the skewness correction, to an interval of fair values,

$$p \to (0.342, 0.658)\,. \tag{9.20}$$

The range (9.20) represents the probability interval for which the outcome interval of the uncertainty bet, for all intents and purposes, is

$$(0, 1000)\,, \tag{9.21}$$

with equality holding at the probabilities

$$p = 0.342, \qquad p = 0.500, \qquad p = 0.658.$$

As the probability p approaches $p = 0.5$ from $p = 0.342$, where (9.21) holds, the skewness interval, in the absence of a kurtosis correction, under shoots the outcome upper bound, with a factor 0.04 of the outcome upper bound[33].

As p crosses the $p = 0.5$ point, the skewness correction transitions from (7.5) to (7.4), and $p = 0.5$, where the skewness is zero, (9.21) holds again.

As p approaches $p = 0.658$, where (9.21) holds, the skewness interval, in the absence of a kurtosis correction, slightly under shoots the outcome lower bound, with a factor 0.04 of the outcome upper bound.

But if we forgo of the skewness interval, and use the sigma interval, and solve the corresponding (9.9) for p. Then, as the initial wealth $m \to \infty$, the fair probability will converge to just the one value,

$$p \to 0.5.$$

So, if it is found that 69% of $N = 68$ subjects prefer decision D_1 over D_2, [33], even though both bets have the same outcome expectation values, then this is because people tend to want to mitigate their losses.

Note that the decision theoretical phenomenon of loss aversion is generally understood to point to the concave down curvature of the Bernoulli law, (4.3). But we have here loss aversion on a meta-level, where, all things being equal, in terms of utility upper and lower bounds, people tend to prefer a possible mitigation of a sure loss.

If we have a certain loss of -250 and an uncertain loss of -1000, then we will be willing to take the uncertain bet, for an initial wealth of $m = 1500$, if the probability of the uncertain event is smaller than

$$p = 0.050.$$

For an initial wealth of $m \to \infty$, this probability converges to

$$p = 0.098. \tag{9.22}$$

If we take such an uncertainty bet, then we adhere to an utility upper bound dominance for losses, which constitutes risk seeking in the negative domain.

9.4. **Risk Aversion II.** The previous analysis may also be performed for the opposite case of a sure gain and a substantial probability of a larger gain. We then

[33] Note that there will also be a commensurate overshoot of the outcome lower bound. But as we know this lower bound to be zero, we already have correct for this overshoot in the confidence bound construction phase.

will see a reversal in the preference for bet D_1 over bet D_2 to a preference for bet D_2 over bet D_1. Risk aversion in the positive domain represents our tendency to secure our profits.

The outcome probability distributions for this problem of choice are[34]:

$$p(O|\, D_1) = \begin{cases} 0.5, & O = 1000 \\ 0.5, & O = 0 \end{cases} \tag{9.23}$$

and

$$p(O|\, D_2) = \begin{cases} 1.0, & O = 500 \end{cases} \tag{9.24}$$

Seeing that this is just another incarnation of Allais' paradox[35], we know that people will tend to prefer bet D_2 over D_1, [4]; and indeed, 80% of $N = 95$ subjects preferred bet D_2 over D_1, [33].

Assuming an initial wealth of $m = 1000$ and by way of (4.3), we find corresponding utility probability distributions:

$$p(u|\, D_1) = \begin{cases} 0.5, & u = q \log \frac{2000}{1000} \\ 0.5, & u = 0 \end{cases} \tag{9.25}$$

and

$$p(u|\, D_2) = \begin{cases} 1.0, & u = q \log \frac{1500}{1000} \end{cases} \tag{9.26}$$

If we solve (9.9) for p, then we find the probability p of the uncertainty bet D_1, for which the bets D_1 and D_2 are undecided:

$$p = 0.764$$

We may plot the fair probability p for a certainty bet as a function of the initial wealth m, where we let $200 < m < 10.000$, Figure 6:

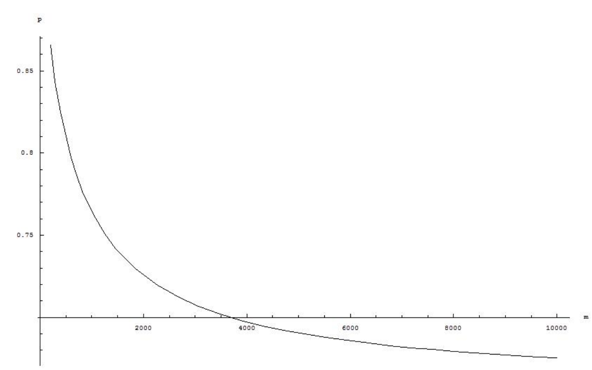

FIGURE 6. Fair probability as function of initial wealth.

---

[34]Compare with (9.15) and (9.16).
[35]See Appendix A.

As the initial wealth $m \to \infty$, the utility of an increment in wealth in the range of $-1000 < \Delta m < -500$ will both become linear and tend to zero, and the fair probability will converge, because of the skewness correction, to an interval of fair values,

$$p \to (0.342, 0.658)\,, \tag{9.27}$$

which is the same interval as (9.20).

So, if it is found that 84% of $N = 70$ subjects prefer decision D_2 over D_1, [33], even though both bets have the same outcome expectation values, then this is because people tend to want to secure their gains.

If we have a certain gain of 250 and an uncertain gain of 1000, then we will beprefer the certain bet, for an initial wealth of $m = 1000$, if the probability of the uncertain event is smaller than

$$p = 0.151.$$

For an initial wealth of $m \to \infty$, this probability converges to, (9.22),

$$p = 0.098.$$

If we take such a certainty bet, then we adhere to an utility lower bound dominance for gains, which constitutes risk aversion in the positive domain.

10. Sixth Supporting Contact: The Psychological Certainty Effect, Part II

In the previous section we defined, for certainty bets, fairness as the decision theoretical equality, (9.9):

$$\left[LB(\,u|\,1, D_2) - LB(\,u|\,p, D_1) = UB(\,u|\,p, D_1) - UB(\,u|\,1, D_2)\right], \tag{10.1}$$

where D_1 and D_2 correspond, respectively, with the uncertainty and certainty bets.

Let O_c and O_u, respectively, be the certainty and the uncertainty outcomes, where $O_c < O_u$. If, for a certainty bet having positive outcomes, we solve (10.1) for the fair probability p, assuming a linear utility for money, we find that the fair probability p maps to the outcome interals

$$(0, 2O_c)\,, \qquad O_c \leq \frac{O_u}{2}, \tag{10.2}$$

which is intuitively fair for the takers of decision D_1, relative to the certainty offer of O_c, and

$$\left[2\left(O_c - \frac{O_u}{2}\right), O_u\right], \qquad O_c > \frac{O_u}{2}, \tag{10.3}$$

which is intuitively fair for the providers of decision D_1, relative to the certainty offer of O_c.

If for an uncertainty pay out of either 0 or $O_u = 5000$, we plot the solution of (10.1) for the fairness probability p, assuming a linear utility for monetary outcomes, as a function of the certainty outcome O_c, we obtain Figure 7:

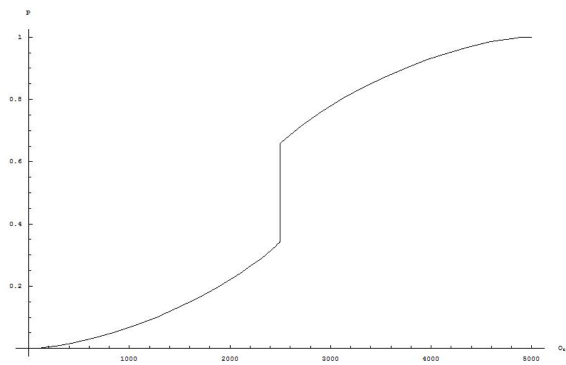

FIGURE 7. Fair probability as function of certain outcome

If we again, but now neglecting the skewness correction, we solve (10.1) for the fairness probability p, assuming a linear utility for monetary outcomes, as a function of the certainty outcome O_c, and add this curve to Figure 7, we obtain Figure 8:

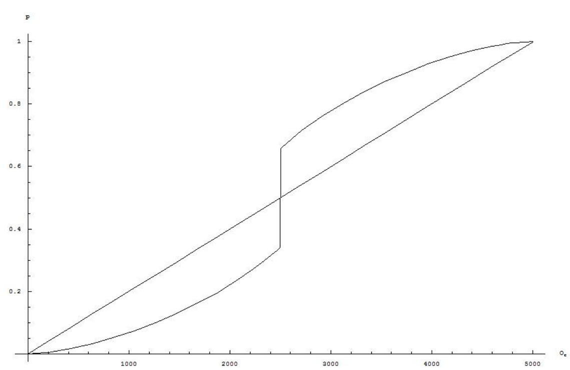

FIGURE 8. Fair probability for sigma and skewness intervals

We may construct, again assuming a linear utility for monetary outcomes, the same graph for the fair probability p of certainty bets involving negative outcomes, where $O_c > O_u$, Figure 9:

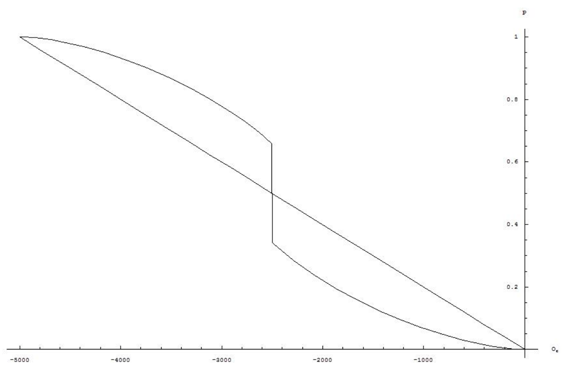

FIGURE 9. Fair probability, sigma and skewness intervals, negative outcomes

If we rescale the x-axes of Figures 8 and 9 as the ratio O_c/O_u, where $|O_c| \leq |O_u|$, and reverse the axes, we obtain the alternative Figure 10:

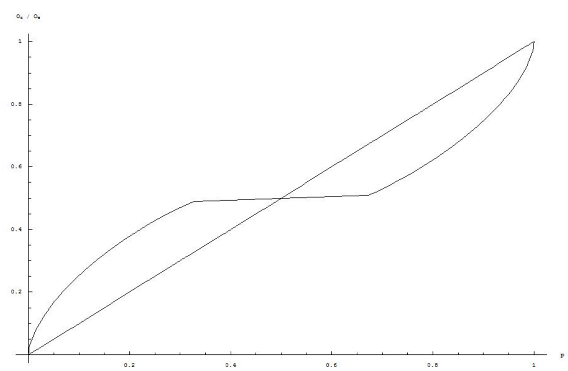

FIGURE 10. Rescaled and rotated figure for sigma and skewness intervals

Now, for those of the readers who are familiar with the cumulative prospect theory, may recognize in Figure 10, Kahneman and Tversky's Figures 1, 2, and 3 of their [60]. Kahneman and Tversky obtained their figures, not from first principles, as we have, but through experimentation, in which subjects where asked to decide on certainty bets of the type we discussed in the previous section.

So, it would seem that Kahneman and Tversky, inadvertently, for they are outspoken anti-Bayesian[36], have provided the Bayesian decision theory with a very strong supporting contact.

Kahneman and Tversky see in the empirical observation of the typical S-curve of Figure 10 a justification for their probability weighing functions,

$$w^{+}(p) = \frac{p^{\gamma}}{p^{\gamma} + (1-p)^{\frac{1}{\gamma}}}, \tag{10.4}$$

[36]See Appendix F.

and

$$w^{-}(p) = \frac{p^{\delta}}{p^{\delta} + (1-p)^{\frac{1}{\delta}}}, \tag{10.5}$$

which over weighs small probabilities and under weighs large probabilities. Moreover, Kahneman and Tversky offer up the implied under weighing of small probabilities, in order to explain the general popularity of lotteries and insurances.

We, on the other hand, see in the empirical observation of the typical S-curve of Figure 10 a confirmation of the intuitive relevancy of the skewness intervals, (7.4) and (7.5).

As we progressed in our research on the Bayesian decision theory, it became obvious to us that the sigma interval, (7.3), though still superior to the 'interval' of expected utility theory, (μ, μ), left out pertinent symmetry information.

All our case studies involved extreme outcomes having small probabilities of occurring, which leaves our probability distributions highly skewed. The presence of skewness leads for sigma intervals to confidence interval coverages which are sub-par, as it will lead to both an under and over shooting of the actual confidence bounds. This is why we felt compelled to search for the skewness interval, (7.4) and (7.5); as this interval promised us more realistic, that is, better informed, criterions of action.

If we drop the assumption of a linear utility of monetary outcomes in the neighborhood of $-5000 < \Delta m < 5000$, and for initial wealths of $m = 1000$ and $m = 6000$ for certainty bets involving, respectively, positive and negative outcomes. Then we may assign, by way of the Bernoulli law, (4.3), utilities to the monetary outcomes. By doing so, we obtain the following fairness ratio outcomes for a given probability probability p of the uncertain proposition, Figures 11 and 12:

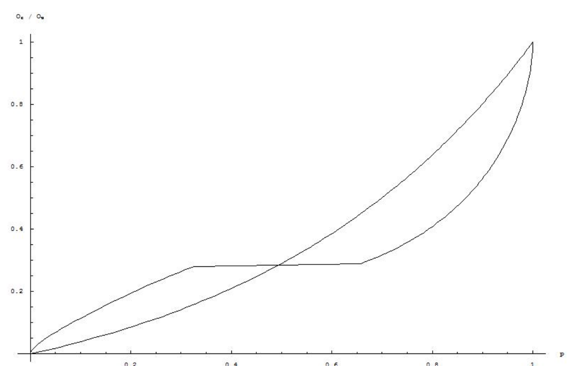

FIGURE 11. Rescaled and rotated figure for positive outcomes

and

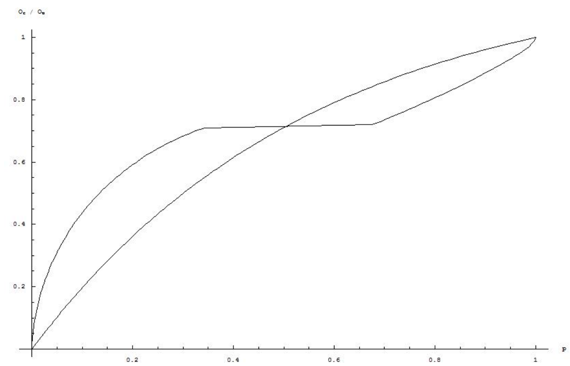

Figure 12. Rescaled and rotated figure for negative outcomes

Comparing Figures 11 and 10, we see that by taking into account the initial wealth m, through the Bernoulli law, (4.3), for low outcome ratios, the fair probabilities p for positive outcomes are adjusted downward, relative to Figure 10. The same holds for large outcome ratios. Furthermore, the fairness symmetry point $p = 0.5$ has been adjusted downward in Figure 11.

Comparing Figures 12 and 10, we see that by taking into account the initial wealth m, through the Bernoulli law, (4.3), for low outcome ratios, the fair probabilities p for negative outcomes are adjusted upward, relative to Figure 10. The same holds for large outcome ratios. Furthermore, the fairness symmetry point $p = 0.5$ has been adjusted upward in Figure 12.

These adjustments make nothing but sense. If we have a small initial wealth, and we stand to gain more than we initially would have gained. Then, for given outcome ratios, we will be more inclined to accept the possibility of gaining nothing, relative to the case where we have a large initial wealth, as the pay-out, in terms of subjective consequences, is relatively larger.

But if we have a small initial wealth, and we stand to lose more than we initially would have lost. Then, for given outcome ratios, we will be less inclined to accept the possibility of losing even more, relative to the case where we have a large initial wealth, as the penalty, in terms of subjective consequences, is relatively larger.

As our initial wealth tends to infinity, and our utility for money becomes linear, we will perceive both problems to be symmetric, as monetary losses are weighed the same as monetary gains, Figure 10.

Furthermore, the differences in the Figures 11 and 12 are commensurate with the fact that Kahneman and Tversky found that their weighing functions for probabilities, (10.4) and (10.5), differed for certainty bets involving positive and negative outcomes.

11. Seventh Supporting Contact: The Ubiquitous Bernoulli Law

We now will give the derivations of the Bernoulli, the Weber-Fechner, and Steven's power laws. It will be seen all that these three laws are equivalent.

11.1. **The Bernoulli Law.**

The utility of a given outcome is the perceived worth of that outcome. If we take the utilities that monetary outcomes hold for us to be an incentive for our decisions, then we may perceive money to be a stimulus.

For the rich man ten dollars is an insignificant amount of money. So, the prospect of gaining or losing 10 dollars will fail to move the rich man, that is, an increment of ten dollars for him has an utility which tends to zero.

For the poor man ten dollars is two days worth of groceries and, thus, a significant amount of money. So, the prospect of gaining or losing ten dollars will most likely move the poor man to action. It follows that an increment of ten dollars for him has an utility significantly greater than zero.

Consider persons A and a, with A having a fortune of 100.000 full-ducats, and with a a fortune of 100.000 semi-ducats, a semi-ducat being the half of a full-ducat. Let f_A and f_a be the moral value functions, defined on, respectively, the monetary full-ducat axis x and the semi-ducat axis $\tilde{x}$. Let x_A and $\tilde{x}_a$ stand for the initial wealths of A and a, respectively; where x_A and $\tilde{x}_a$ are points on the monetary axes x and $\tilde{x}$, respectively.

Bernoulli derived his law by way of three simple symmetry considerations for the moral functions f_A and f_a, [6, 52]:

(1) For an arbitrary increment c in wealth, the moral movement of this increment will be less for the rich man, than for the poor man; that is, if we make for f_a the appropriate change of variable, from $\tilde{x}$ to x, then we have that

$$\left.\frac{d}{dx} f_A(x)\right|_c < \left.\frac{d}{dx} f_a(x)\right|_c .$$

From which it follows that effect of c on a given f decreases as the initial wealth increases.

(2) It is proposed that the movement in a general moral value function f, for a given positive increment dx, is proportional to the value of this increment; that is,

$$\left.\frac{d}{du} f(u)\right|_{c=dx} \propto dx,$$

as this is the simplest function for which f increases as a function of an increment in x.

(3) Furthermore, it is proposed that this movement in f is inversely proportional to the value of the initial wealth x; that is,

$$\left.\frac{d}{du} f(u)\right|_{c=dx} \propto \frac{1}{x}.$$

where '$\propto$' is the proportionality sign.

Bernoulli arrived at his third consideration, using the following reasoning. The change in moral value of c full-ducats for A will be half the change in moral value of c full-ducats for a. Only if either a sees his fortune increased to 100.000 semi-ducats, or, equivalently, 100.000 full-ducats, or if A sees his fortune reduced to 50.000 full-ducats, or, equivalently, 100.000 semi-ducats, only then will a have the same change in moral value as A for c full-ducats.

We then have that, if we make for f_a the appropriate change of variable from $\tilde{x}$ to x,

$$\frac{\left.\frac{d}{dx} f_A(x)\right|_c}{\left.\frac{d}{dx} f_a(x)\right|_c} = \frac{x_a}{x_A}, \tag{11.1}$$

where x_a is the initial fortune of a, translated from the semi-ducat $\tilde{x}$-axis to the full-ducat x-axis.

It follows from (11.1) that we have, in general, that the change in moral value is inversely proportional to the initial we hold, that is,

$$\left.\frac{d}{dx} f(x)\right|_c \propto \frac{1}{x}, \tag{11.2}$$

which is Bernoulli's third consideration.

If we combine the second and the third consideration, we obtain the differential equation

$$f'(x) = q\frac{dx}{x}, \tag{11.3}$$

which, if solved for the boundary condition that for a given person with an initial wealth of x_0 an increment of zero holds no utility, either negative or positive, gives

$$f(x) = q \log \frac{x}{x_0}, \tag{11.4}$$

which may be rewritten as

$$f(\Delta x | x_0) = q \log \frac{x_0 + \Delta x}{x_0}. \tag{11.5}$$

11.2. **The Weber-Fechner Law.** Let S signify stimuli intensity and let Q signify sensation strength. Weber's law states that the increment ΔS needed to elicit a judgment that $S + \Delta S$ is just noticeably different from S is proportional to S:

$$\Delta S = wS, \tag{11.6}$$

where w is a positive constant dependent upon the specific type of sensory stimulus offered and ΔS is understood to be the stimulus increment corresponding with a just noticeable difference.

Fechner generalized the experimental Weber law by stating that all differences in sensational strength, and not only the ones that are just noticeable, are proportional to the relative change $\Delta S/S$, that is,

$$\Delta Q = q\frac{\Delta S}{S}. \tag{11.7}$$

where k is a positive constant dependent upon the specific type of sensory stimulus offered and ΔS is now understood to be the stimulus increment corresponding with the increment in sensation strength ΔQ.

Dividing both sides of (11.7) by ΔS gives

$$\frac{\Delta Q}{\Delta S} = q\frac{1}{S}. \tag{11.8}$$

Fechner then makes the assumption that, just as a physically small quantity ΔS can be reduced without limit to the differential dS, so a small quantity of sensation can be reduced without limit to the differential dQ. By way of this assumption, we may let (11.8) tend to the differential equation

$$\frac{dQ}{dS} = q\frac{1}{S}. \tag{11.9}$$

The general solution of this differential equation is

$$Q = q\log S + c, \tag{11.10}$$

where c is some constant of integration.

Introducing an initial value condition for (11.9) that says that at stimulus value S_0 there is no sensation strength, that is, $Q(S_0) = 0$, leaves us with the Weber-Fechner law

$$Q\left(S|\, S_0\right) = q\log\frac{S}{S_0}, \tag{11.11}$$

or, equivalently,

$$Q\left(\Delta S|\, S_0\right) = q\log\frac{S_0 + \Delta S}{S_0}. \tag{11.12}$$

The Weber-Fechner law, (11.12), is identical to the utility function which had been proposed a century earlier by Bernoulli, (11.5).

Fechner himself was aware of this equivalence. Nonetheless, he believed his derivation to be the more general. Fechner argued that Bernoulli's derivation only applied to the special case of utility, whereas his law, though identical, applied to all sensations, as it invokes Weber's law.

However, as pointed out in [52], Fechner failed to provide any compelling reason why the principles employed in Bernoulli's derivation of the subjective value of objective monies should not be extendible to sensations in general. Nontheless, we do believe that Fechner was of good faith, in denying Bernoulli scientific primacy.

First of all, Fechner called the Weber-Fechner law, when he first published it, the Weber law. Second of all, Fechner had a deep spiritual need for some kind of harmony between the physical and mental universes, and the Weber-Fechner law provided him with this harmony, for this law spoke of the basic oneness of the physical and mental universes

The Weber-Fechner law demonstrated that both universes adhered to seemingly mechanistic laws. It then followed that the freedom of the latter universe, in terms of free will and volition, implied, by way of analogy, a commensurate freedom of the former; thus, opening the way for the possibility of a besouled physical universe. Which had become Fechner's only hope for spiritual salvation, [15].

We can imagine that Fechner might have felt that a law that assigned subjective values to objective monies was too arbitrary and sordid a foundation for the lofty purpose he wished it to serve. In contrast, the initial Weber law allowed Fechner to forgo of the money argument and derive a law, which though in form identical to Bernoulli's, differed in that it applied to all human sensations.

11.3. **Steven's Power Law.** Steven's power law is based on the observation, that it is the ratio $\Delta Q/Q$, rather than the difference ΔQ, that is proportional to $\Delta S/S$, [57]. This observation leads to the equality

$$\frac{\Delta Q}{Q} = q\frac{\Delta S}{S}. \tag{11.13}$$

Letting the differences in Q and S go to differentials, we may rewrite (11.13) as

$$\frac{dQ}{Q} = q\frac{dS}{S}. \tag{11.14}$$

This equation has its general solution

$$\log Q = q \log S + c'. \tag{11.15}$$

Taking the exponent of both sides of (11.15), we get the power law for stimulus perception

$$Q = cS^q, \tag{11.16}$$

where $c = \exp(c')$.

Stevens found the power law to hold for several sensations; binaural and monaural loudness, brightness, lightness, smell, taste, temperature, vibration duration, repetition rate, finger span, pressure on palm, heaviness, force of hand grip, autophonic response, and electric shock, [57].

The power law is applied by letting subjects compare the sensation ratio of Q_1 to Q_0 for corresponding stimuli strengths S_1 and S_0:

$$\frac{Q_1}{Q_0} = \left(\frac{S_1}{S_0}\right)^q. \tag{11.17}$$

Let $S_1 = S_0 + \Delta S$, where ΔS is some increment, then we may rewrite (11.17) as

$$\frac{Q_1}{Q_0} = \left(\frac{S_0 + \Delta S}{S_0}\right)^q. \tag{11.18}$$

For an increment of $\Delta S = 0$, the ratio of perception stimuli will be $Q_1/Q_0 = 1$. Taking the log of the ratio (11.18) we may map the ratio of perceived stimuli to a corresponding utility scale where a zero increment ΔS corresponds with a zero utility:

$$Q'\left(\left.\Delta S\right| S_0\right) = \log\frac{Q_1}{Q_0} = q \log\frac{S_0 + \Delta S}{S_0} \tag{11.19}$$

But this is just the Weber-Fechner law, (11.13).

11.4. **Summary.** The Weber-Fechner law gives us just noticeable differences on a log scale, (11.13). The power law gives us ratios of sensation strengths, (11.18). Taking the log of the ratio of sensation strengths, we may obtain the just noticeable differences again, (11.19). But the Weber-Fechner for just noticeable differences is just the Bernoulli law for utilities, (11.5).

We refer the reader to [52], for a discussion of Thurnstone's derivation of the satisfaction law. This law, which takes as its input the increment in the number of items of commodity, is also of the form of Bernoulli's law.

12. Eight Supporting Contact: The Negative Bernoulli Law

In this section we present the negative Bernoulli law for debts, which is a corollary of the Bernoulli law for income.

The negative Bernoulli law predicts that for the very poor, having a small initial wealth and large initial debts, a large loss of direct income will be more devastating, than an increase of, say, twice that loss in their long-term debt. This law also explains why, for these poor, having a small initial wealth and large initial debts, the temptation to take out loans, if offered the opportunity, will be quite great, [30].

In this section we will discuss the Bernoulli law and its scaling constant q in a psycho-physical setting, which is why we sometimes will refer to them as, respectively, the Weber-Fechner law and the Weber constant.

12.1. **The positive Bernoulli law.** The translation of monetary stimuli to utilities is analogous to the case where we are asked to translate loudness to a numerical value. According to Weber-Fechner law, postulated in the 19th century[37] by the

[37] Which is just Bernoulli's law, which he postulated in the 18th century.

experimental psychologist Fechner, intuitive human sensations tend to be logarithmic functions of the difference in stimulus, [16]. So, we do not perceive stimuli in isolation, rather we perceive the relative change in stimuli, case in point being the decibel scale of sound.

Let S_1 and S_2 be two stimuli which are to be compared. Then the Weber-Fechner law tells us that the Relative Change (RC) is the difference of the logarithms of the stimuli:

$$\text{RC} = c \log_d S_2 - c \log_d S_1 = c \log_d \frac{S_2}{S_1}, \tag{12.1}$$

where c is some scaling factor and d some base of the logarithm. From (12.1), we have that if stimuli S_1 and S_2 are indistinguishable, that is, of the same strength, then their RC is 0. If S_2 increases relative to S_1, then RC > 0. If S_2 decreases relative to S_1, then RC < 0.

The Weber-Fechner law allows for one degree of freedom. This can be seen as follows. Since

$$\log_d x = \frac{\log x}{\log d},$$

we can rewrite (12.1) as

$$\text{RC} = q \log \frac{S_2}{S_1}, \tag{12.2}$$

where

$$q = \frac{c}{\log d}. \tag{12.3}$$

Let ΔS be an increment, either positive or negative, in a monetary stimulus S. Then we may define the utility of a monetary increment ΔS to be the perceived relative change in the initial wealth S due to that increment ΔS, (12.2):

$$u(\Delta S|\, S) = q \log \frac{S + \Delta S}{S}, \qquad -S < \Delta S < \infty. \tag{12.4}$$

If $\Delta S = -S$, then (12.4) tells us that a loss of all one's initial wealth S would have a utility of minus infinity. This is clearly not realistic. So, in order to model such a loss, we must introduce the threshold of income which is still significant γ, [28], where $\gamma > 0$. The threshold of income has the following interpretation. .

Even for a beggar there is some minimum amount of money that is still significant. This may be one dollar for a bag of potato chips, or three dollars for a packet of cigarettes. If the loss of money breaks through the limit of the minimum significant amount γ, the beggar is left with an amount of money which, for all intents and purposes, is worthless.

Using the concept of the threshold of income, we may modify (12.4) as

$$u(\Delta S|\, S) = q \log \frac{S + \Delta S}{S}, \qquad -S + \gamma < \Delta S < \infty. \tag{12.5}$$

If we want to give a graphical representation of (12.5), then the scaling constant q, also known as the Weber constant, must be set to some numerical value.

Say, we have a monthly expendable income of a thousand dollars, for groceries and the like, then introspection[38] would suggest that a loss or gain of an amount less than ten dollars would not move us that much.

So, $\Delta S = 10$ constitutes a just noticeable difference, or, equivalently, 1 utile, for an initial wealth of $S = 1000$, (12.5):

$$1 \text{ utile} = q \log \frac{1000 + 10}{1000}. \tag{12.6}$$

If we then solve for the unknown Weber constant q, we find

$$q = \frac{1}{\log 1010 - \log 1000} \approx 100. \tag{12.7}$$

Note that utiles represent the utility of the monetary outcomes, much like decibels represent the perceived intensity of sound[39].

Suppose we have a student who has three hundred dollars per month to spend on groceries and the like and who stands to lose or to gain up to two hundred dollars. Then, by way of (12.5) and (12.7), we obtain the following mapping of monetary outcomes to utilities, Figure 13:

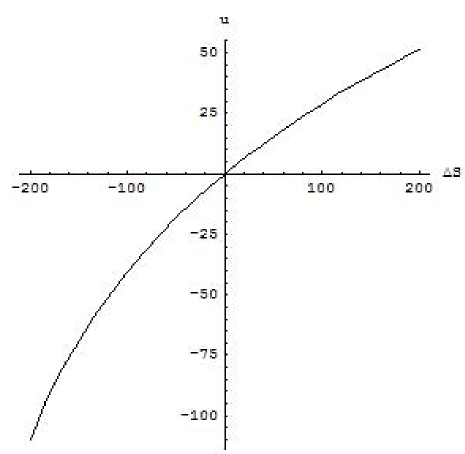

Figure 13. Utility plot for initial wealth 300 dollars

For the case of the rich man who has one million dollars to spend on groceries and the like and who stands stands to lose or to gain up to a hundred thousand dollars, we obtain the alternative mapping, Figure 14:

[38] Introspection being the starting point of all psychological experimentation.

[39] Note that for the decibel scale the Weber constant has been determined to be $q = 10/\log 10 = 4.34$.

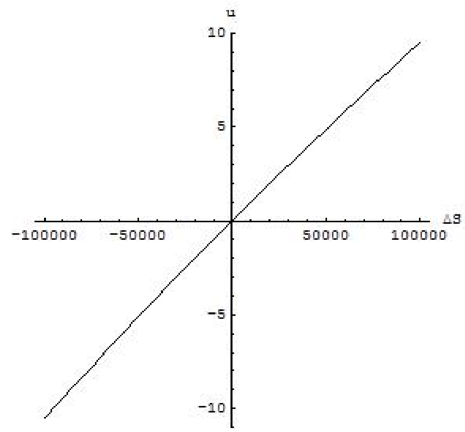

FIGURE 14. Utility plot for initial wealth 1.000.000 dollars

Loss aversion is the phenomenon that losses may loom larger than gains, [60]. Comparing Figures 13 and 14, we see that the Weber-Fechner law of experimental psychology captures both the loss aversion of the poor student, that is, asymmetry in gains and losses, as well as the linearity of the utility of relatively small gains and losses for the rich man.

12.2. **The negative Bernoulli law.** Until now we have treated only the case were the maximal loss did not exceed the initial wealth m. However, in real life we may lose more than we actually have, by way of debt. So, we now proceed to assign utilities to increments in debt.

According to the Weber-Fechner law we cannot lose more money than we initially had. Otherwise we may have that the ratio in the logarithm in the Weber-Fechner-Law, (12.4),

$$u(\,\Delta S|\,S) = q\log\frac{S+\Delta S}{S}, \tag{12.8}$$

may become negative, leading to a breakdown of the logarithm.

However, whenever we incur a debt we lose more money than we have. Furthermore, we can have a debt and an income, both at same time. So, we propose that there are two different monetary stimuli dimensions in play; the first dimension being an actual income dimension and the second dimension being a debt dimension.

We propose to model the debt utilities by way of the negative Weber-Fechner law:

$$u(\,\Delta D|\,D) = -b\log\frac{D+\Delta D}{D}, \tag{12.9}$$

where we let D be the initial debt, ΔD the increment in debt, and b the the Weber constant of a monetary debt.

The rationale behind (12.9) is as follows. If we view a debt increment as a stimulus, then it follows that we may use the psycho-physical Weber-Fechner law in the determination of the moral value of a given debt increment.

For positive increments ΔD, there is an increase in current debt, whereas for negative increments ΔD, there is a decrease in current debt. In order to assign both a negative utility to an increase in current debt and positive utility to a decrease in debt, we need to multiply the Weber-Fechner law times minus one, (12.9).

If we have no initial debt, that is, $D = 0$, then (12.9) tells us that any positive increment in debt ΔD would have a utility of minus infinity. This is clearly not realistic. So, in order to model an increment in debt for those who are without debt, we must introduce a minimum significant amount of debt which is equal to minimum significant amount of income, γ.

The threshold amount of debt, γ, may also be used in the case of $\Delta D = -D$, in order to prevent an infinite utility being assigned to a full repaying of one's debts. Using the concept of the minimum significant amount of debt stimulus, we may modify (12.9) as

$$u(\,\Delta D|\, D) = -b \log \frac{D + \Delta D}{D}, \qquad -D + \gamma < \Delta D < \infty, \tag{12.10}$$

If we want to give a graphical representation of (12.10), then the Weber constant b, must be set to some numerical value.

Say, we have a total debt of forty thousand dollars, in the form of a student loan, which we eventually will have to pay back, but not right now. Then introspection would suggest that a increment or decrement of an amount less than a thousand dollars would not move us that much.

So, $\Delta D = 1000$ constitutes one utile, or, equivalently, a just noticeable difference in debt for an initial debt of $D = 40.000$, that is, (12.5):

$$1 \text{ utile} = -b \log \frac{40.000 - 1000}{40.000}. \tag{12.11}$$

If we then solve for the unknown Weber constant b of debt stimuli,

$$b = -\frac{1}{\log 390000 - \log 40000} \approx 40, \tag{12.12}$$

we find this Weber constant to be smaller by a factor of 2.5 than the Weber constant q of income stimuli, (12.7).

It is well possible that this difference in Weber constants can be attributed to the difference in abstractness of the concepts. The losing of actual monies is quite concrete, whereas the accrueing of a debt, repayable somewhere in a distant future, is somewhat more abstract.

But there is always a chance that these authors were off in their introspection[40], and that both Weber constants should be approximately equal. We leave this issue, together with the psychological reality of the phenomenon of debt relief, given below, for future psychological experimentation, as we proceed with our discussion of the debt utilities.

Suppose that a student has a student loan which has accumulated to forty thousand dollars. Then, by way of (12.9) and (12.12), we obtain the following mapping of increments in debt to utilities, Figure 15.

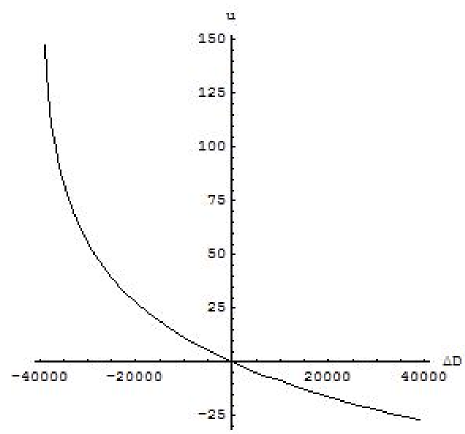

Figure 15. Utility plot for initial debt 40.000 dollars

As stated previously, loss aversion is the phenomenon that losses may loom larger than gains. In Figure 15 we see the phenomenon that debt reduction may loom larger than debt increase. We will call this corollary of the psycho-physical Weber-Fechner law: 'debt relief', the relief of loosing one's debts.

Now, does the phenomenon of debt relief correspond with a real psychological phenomenon? We belief that it actually does.

Say, we have a debt of a a thousand dollars. Then we can imagine ourselves feeling greatly reliefed, were we to be released of our debt. Now, were our debt, instead, to be doubled to two thousand dollars, then we can also imagine ourselves feeling unhappy about this. But this feeling of unhappiness about the doubling of our debt would be of a lesser intensity than the corresponding relief of having our debt acquitted.

[40]Note that actual value of the Weber constants q and b of, respectively, income and debt stimuli have no direct bearing on any of the results given in this fact sheet; save the handful of examples which are given in this section, in order to demonstrate the qualitative behavior of the negative Weber-Fechner law, or, equivalently, the negative Weber-Fechner law.

We will now look at the practical implications of the negative Bernoulli law, (12.9), and its Weber constant b, (12.12).

A student loan initially represents a gain in debt stimulus. This debt makes itself felt, in terms of actual loss of income, only after graduation, the moment the monthly payments have to be paid and take a considerable chunk out of one's actual income.

Say, that the student of Figure 15, having become a PhD, and having a net income of fifteen hundred dollars, is called upon to make good on his loan, by way of monthly payments of five hundred dollars. Then these payments represent both a loss in income, having a negative utility of, (12.8) and (12.7):

$$u^{(\text{income-loss})} = 100 \log \frac{1500 - 500}{1500} = -41.5, \tag{12.13}$$

as well as a decrements in debt, having a positive utility of, (12.9) and (12.12):

$$u^{(\text{debt-decrease})} = -40 \log \frac{40000 - 500}{40000} = 0.5, \tag{12.14}$$

It follows that our PhD can find little to no comfort in the fact that he is paying of his debt, as he acutely feels the sting of loss of income. This is, together with the difference in Weber constants[41], (12.7) and (12.12), reflective of the fact that his utility function for income is highly non-linear in the neighborhood of the increment, whereas his utility function for debt is highly linear in that region.

Now, say that we have another PhD, who during his student days lived a more frugal life style and, consequently, only has a debt of two thousand dollars. For this PhD student, when called upon to make good on the loan, the loss of income will be felt just as keenly, with a negative utility of $u = -40.5$, (12.13). However, he will find more satisfaction in the fact that he is paying of his debts, (12.9) and (12.12):

$$u^{(\text{debt-decrease})} = -40 \log \frac{2000 - 500}{2000} = 11.5, \tag{12.15}$$

seeing that he has a more curved utility function for debt than our previous PhD student.

Nontheless, the first PhD student may feel, after a couple of years of monthly repayments, when his loan has been reduced to twenty thousand dollars, for the first time, as if he has an actual stake in the repayment of his debt, (12.9) and (12.12):

$$u^{(\text{debt-decrease})} = -40 \log \frac{20000 - 500}{20000} = 1.0, \tag{12.16}$$

as his debt repayment utility crosses the threshold of the just noticeable difference.

The negative Bernoulli law also gives an explanation why for the very poor, having a minimum monthly wage of seven hundred euros, and already having a

[41]A difference which accounts only for a factor of 2.5 in the observed differences of the utilities (12.13) and (12.14).

large debt of, say, twenty thousand euros, a loss of income of, say, five hundred euros, is perceived to be so much more devastating than an increase in debt of, say, a thousand euros.

For this poor person, the loss of actual income has a negative utility of -125 utiles and the gain of an increase of has a negative utility of only -2 utiles[42].

Likewise, the temptation for the very poor, if offered the opportunity, to take out a loan of a thousand euros will be quite great.

As for this poor person, the immediate gain of a direct increase of a thousand euros in income will have a positive utility of $+89$ utiles, whereas the negative utility of an increase in debt of a thousand dollars will have a negative utility of only -2 utiles[43], [30].

13. A Consistency Proof of the Bernoulli Law

In the Bayesian decision theory, we start by constructing our outcome probability distributions, by way of the product, sum, and generalized rules of the Bayesian probability theory[44]. We then proceed to assign utilities to the outcomes of these probability distributions, by way of the Bernoulli law, in order to construct our utility probability distributions. Finally, we compare the location of these utility probability distributions by way of some function of the cumulants of the utility probability distributions.

The product, sum, and generalized rules of the Bayesian probability theory are the only consistent operators on probabilities, [8, 28, 39]. So, consistency wise, we have no choice but to use these rules to construct our outcome probability distributions.

The cumulant function initially proposed by Bernoulli was the identity function for the first cumulant, that is, the expectation value of the utility probability distribution. But this proposal, though sufficient enough in many cases, may, nonetheless, lead to Ellsberg and Allais 'paradoxes', which is an indication that the information of the higher order cumulants should also be taken into account[45].

If we take as our function of the cumulants of the utility probability distribution the skewness intervals, (7.4) and (7.4), then we find that all these paradoxes fall away and, moreover, leave us with a decision theoretical algorithm, which is both surprisingly rich in structure and eminently intuitive.

In order to map monetary outcomes of the outcome probability distributions to their corresponding utilities, and so construct the utility probability distributions, we make use the Bernoulli law.

[42] Even if $b = 100$, (12.12), this negative utility will only be -5 utiles.
[43] Idem.
[44] See Appendix E.
[45] See Appendix A.

This law is the one remaining degree in the Bayesian decision theory. In this section we will give the derivation of the Bernoulli law, by way of consistency constraints on the lattice of ordering.

13.1. **Lattice Theory and Quantification.** Two elements of a set are ordered by comparing them according to a binary ordering relation, that is, by way of '$\leq$', which may be read as 'is included by'. Elements may be comparable, in which case they form a chain, or they may be incomparable, in which case they form a antichain. A set consisting of both inclusion and incomparability are called partially ordered sets, or posets for short, [36].

Given a set of elements in a poset, their upper bound is the set of elements that contain them. Given a pair of elements x and y, the least element of the upper bound is called the join, denoted[46] $x \vee y$. The lower bound of a pair of elements is defined dually by considering all the elements that the pair of elements share. The greatest elements of the lower bound is called the meet, denoted $x \wedge y$.

A lattice is a partially ordered set where each pair of elements has a unique meet and unique join. There often exist elements that are not formed from the join of any pair of elements. These elements are called join-irreducible elements. Meet-irreducible elements are defined similarly. We can choose to view and join and meet as algebraic operations that take any two lattices elements to a unique third lattice element. From this perspective, the lattice is an algebra.

An algebra can be extended to a calculus by defining functions that take lattice elements to real numbers. This enables one to quantify the relationships between the lattice elements.

A valuation v is a function that takes a single lattice element x to a real number $v(x)$ in a way that respects the partial order, so that, depending on the type of algebra, either $v(x) \leq v(y)$ or $v(y) \leq v(x)$, if in the poset we have that $x \leq y$. This means that the lattice structure imposes constraints on the valuation assignments, which can be expressed as a set of constraint equations, [38].

The Ordering Space. The set of all possible orderings is called the ordering space.

The lattice of ordering is generated by taking the power set, which is the set of all possible subsets of the set of all order elements, say, x, y, z, etc..., where $x < y < z <$ etc..., and ordering them according to Polya's min-max rule, [37], where the meet $\wedge$ is defined as

$$x \wedge y = \min x, y = x, \tag{13.1}$$

[46]Note that we over-load the symbol '$\vee$' here, which still stands for disjunction, though now in the general context of lattice theory.

and the join is defined as

$$x \vee y = \max x, y = x. \tag{13.2}$$

The ordering relation of the min-max rule naturally encodes ordering, such that an ordering element higher up on the lattice is always greater or equal than all the connecting elements below it. Likewise, an ordering element further down on the lattice is always smaller or equal than all the connecting elements below it

For example, y is greater than the lower lattice element $x \vee y$, $x \vee y \vee z$, to which it is directly connected, by way of $x \vee y$ and $y \vee z$. Likewise, y is greater than the higher lattice element $x \wedge y$, $x \wedge y \vee z$, to which it is directly connected, by way of $x \wedge y$ and $y \wedge z$. In this sense the lattice of ordering is an algebra.

In what follows we derive a measure, called the Bernoulli law, that quantifies the degree of ordering.

13.2. **The general sum rule.** We begin by considering a special case of elements x and y with join $x \vee y$. In Figure 16 we give the graphical representation of this simple lattice.

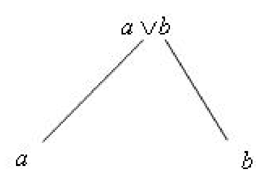

FIGURE 16. Lattice of $x \vee y$

The value we assign to the join $x \vee y$, written $v(x \vee y)$, must be a function of the values we assign to both x and y, $v(x)$ and $v(y)$. Since, if there did not exist any functional relationship, then the valuation could not possibly reflect the underlying lattice structure; that is, valuation must maintain ordering, in the sense that $x \leq x \vee y$ implies either $v(x) \leq v(x \vee y)$ or $v(x) \geq v(x \vee y)$.

So, we write this functional relationship in Figure 16 in terms of an unknown binary operator $\oplus$:

$$v(x \vee y) = v(x) \oplus v\left(y\right). \tag{13.3}$$

We now consider another case where we have three elements x, y, and z, Figure 17.

Because of the *associativity* of the join, we have that the least upper bound of these three elements, $x \vee y \vee z$, can be obtained in these two different ways:

$$x \vee (y \vee z) \qquad \text{and} \qquad (x \vee y) \vee z. \tag{13.4}$$

Figure 17. Lattice of $x \vee y \vee z$

By applying (13.3) to (13.4), the value we assign to this join can also be obtained in two different ways:

$$v(x) \quad [v(y) \quad v(z)] \qquad \text{and} \qquad [v(x) \quad v(y)] \quad v(z): \tag{13.5}$$

Consistency then demands that the equivalent assignments (13.5) have the same value:

$$v(x) \quad [v(y) \quad v(z)] = [v(x) \quad v(y)] \quad v(z): \tag{13.6}$$

This the functional equation for the operator , for which the general solution is given by, [1]:

$$f[v(x \vee y)] = f[v(x)] + f[v(y)]; \tag{13.7}$$

where f is an arbitrary invertible function, so that many valuations are possible. We define the valuation u as

$$u(x) \quad f[v(x)];$$

and rewrite (13.7) as

$$u(x \vee y) = u(x) + u(y): \tag{13.8}$$

Now that we have a constraint on the valuation for our simple example, we seek the general solution for the entire lattice. To derive the general case, we consider the lattice in Figure 18.

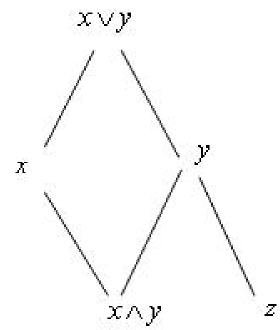

FIGURE 18. Extended lattice of $x \vee y$

If we apply (13.8) to both the elements y and $x \vee y$, we get

$$u\left(y\right) = u\left(x \wedge y\right) + u\left(z\right), \tag{13.9}$$

and, since y is just the join of the part it shares with x joined with z, where, for the lattice of ordering, z is understood to be the meet, (13.1), with y and some other ordering element to the right,

$$u\left(x \vee y\right) = u\left(x\right) + u\left(z\right). \tag{13.10}$$

Substituting for $u\left(z\right)$ in (13.9) and in (13.10), we get the general sum rule:

$$u\left(x \vee y\right) = u\left(x\right) + u\left(y\right) - u\left(x \wedge y\right) \tag{13.11}$$

In general, for bi-valuations we have

$$w\left(\left. x \vee y\right| t\right) = w\left(\left. x\right| t\right) + w\left(\left. y\right| t\right) - w\left(\left. x \wedge y\right| t\right), \tag{13.12}$$

for any context t, [39].

Note that the sum rule is not focused solely on joins since it is symmetric with respect to interchange of joins and meets.

At this point we have *derived* additivity of the measure, which is considered to be an axiom of measure theory. This is significant in that associativity constrains us to have additive measures - there is no other option, [39].

If we apply (13.1) and (13.2), which are the operators of this particular lattice, to (13.12), we are left with the platitude

$$w\left(\left. y\right| t\right) = w\left(\left. x\right| t\right) + w\left(\left. y\right| t\right) - w\left(\left. x\right| t\right) = w\left(\left. y\right| t\right), \tag{13.13}$$

which, nonetheless, is very consistent.

So, we find that on the lattice of ordering the general sum rule provides no other constraint than that the quantification w should assign the same value to the same argument, which we intended to do anyway.

Now, for both the lattice of statements and questions, which quantify, respectively, to the Bayesian probability and information theories, we have that the general sum rule (13.12) is a highly non-trivial operator, as it gives rise to the general sum rule of the measures of probability and relevancy, respectively.

Chain Rule. We now focus on bi-valuations and explore changes in context [38]. We begin with a special case and consider four ordered elements $x \leq y \leq z \leq t$.

The relationship $x \leq z$ can be divided into the two relations $x \leq y$ and $y \leq z$. In the event that z is considered to be the context, this sub-division implies that the context can be considered in parts. The bi-valuation we assign to x with respect to the context z, that is, $w\left(x|\, z\right)$, must be related to both the bi-valuation we assign to x with respect to the context y, that is, $w\left(x|\, y\right)$, and the bi-valuation we assign to y with respect to the context z, that is, $w\left(y|\, z\right)$.

So, there exists a binary operator $\otimes$ that relates the bi-valuations assigned to the two steps to the bi-valuation assigned to the one step:

$$w\left(x|\, z\right) = w\left(x|\, y\right) \otimes w\left(y|\, z\right). \tag{13.14}$$

By extending (13.14) to three steps, and considering the bi-valuation $w\left(x|\, t\right)$, relating x and t, via intermediate contexts y and z, we get Figure 19.

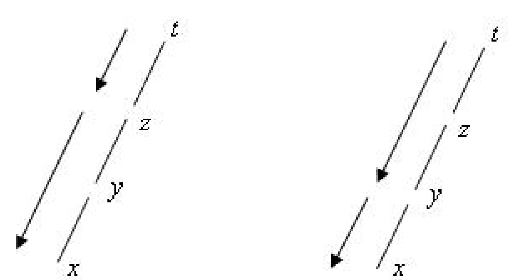

FIGURE 19. Context lattice of t

This figure leads to the associativity relationship:

$$\left[w\left(x|\, y\right) \otimes w\left(y|\, z\right)\right] \otimes w\left(z|\, t\right) = w\left(x|\, y\right) \otimes \left[w\left(y|\, z\right) \otimes w\left(z|\, t\right)\right]. \tag{13.15}$$

By way the associativity theorem, [39], we have that any operator, be it $\oplus$, $\otimes$, or $\odot$, has a scale on which associativity relations (13.6) and (13.15) are additive, which would seem to solve our (13.15) trivially.

However, once we have fixed the behavior of w to be additive with respect to either the arguments before the solidus or the arguments behind the solidus, we can not regrade to that scale anymore. We then will have to infer additivity on some other grade, say, $\Theta(w)$, [39].

For example, in the quantification of the lattice of statements we are forced to infer additivity on the alternative grade Θ; seeing that we have lost the degree of freedom of addition on the grade w when we find a non-trivial generalized sum rule (13.15), [39].

So, for the lattice of statements, we have that the chain rule for context change forces us to use addition on the alternative grade Θ, which leaves us with the equality:

$$\Theta\left[w\left(x|\, z\right)\right] = \Theta\left[w\left(x|\, y\right)\right] + \Theta\left[w\left(x|\, y\right)\right], \tag{13.16}$$

If we solve (13.16), it is found that Θ is the logarithmic function times some constant q, which we may set to one, if we so like, [39].

Since we have that the inverse of the logarithmic function is the exponential function, we may label this inverse as, say, Ψ. We then return to our original grade w, on which we have derived the general sum rule, by way of inversion:

$$\begin{aligned} w\left(x|\, z\right) &= \Psi\left\{\Theta\left[w\left(x|\, y\right)\right] + \Theta\left[w\left(x|\, y\right)\right]\right\} \\ &= e^{\log\left[w\left(x|y\right) w\left(x|y\right)\right]} \\ &= w\left(x|\, y\right) w\left(x|\, y\right) \end{aligned} \tag{13.17}$$

which is the product rule of Bayesian probability theory.

Now, seeing that for the Bernoulli law addition on the w grade is still allowed, (13.13), we have that by way of the associativity theorem, (13.15) results in a constraint equation for non-negative bi-valuations involving changes in context [39]:

$$w\left(x|\, z\right) = w\left(x|\, y\right) + w\left(y|\, z\right), \tag{13.18}$$

where the grade w in (13.18) is the same as the grade $\Theta(w)$ in (13.16).

It then follows, by way of [39], that $w(x)$ is of the form $q\log(x)$, which leaves with the change of context rule of decreasing orderings

$$q_d \log\left(x|\, z\right) = q_d \log\left(x|\, y\right) + q_d \log\left(y|\, z\right). \tag{13.19}$$

where q_d is the chain rule constant for decreasing orderings.

Alternatively, if x is considered to be the context, rather then z, then the subdivision of $x \leq z$ in relations $x \leq y$ and $y \leq z$ also implies that the context can be considered in parts.

The bi-valuation we assign to z with respect to the context x, that is, $w\left(z|\, x\right)$, must be related to both the bi-valuation we assign to z with respect to the context y, that is, $w\left(z|\, y\right)$, and the bi-valuation we assign to y with respect to the context x, that is, $w\left(y|\, x\right)$.

So, there again exists a binary operator $\otimes$ that relates the bi-valuations assigned to the two steps to the bi-valuation assigned to the one step

$$w\left(z|\, x\right) = w\left(z|\, y\right) \otimes w\left(y|\, x\right). \tag{13.20}$$

By extending (13.20) to three steps, and considering the bi-valuation $w\left(t|\, x\right)$, relating t and x, via intermediate contexts z and y, we get Figure 20.

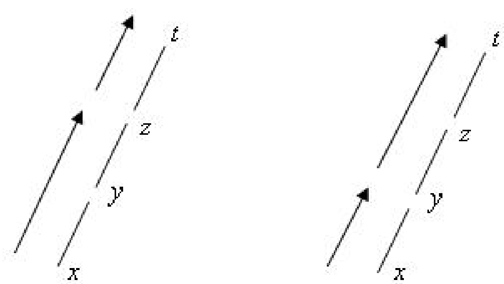

Figure 20. Context lattice of t

This figure leads to the associativity relationship:

$$w\left(t|\, z\right) \otimes \left[w\left(z|\, y\right) \otimes w\left(y|\, x\right)\right] = \left[w\left(t|\, z\right) \otimes w\left(z|\, y\right)\right] \otimes w\left(y|\, x\right). \tag{13.21}$$

This relationship then leads us, by way of (13.16), (13.17), and (13.18), to the change of context rule of increasing orderings

$$q_i \log\left(z|\, x\right) = q_i \log\left(z|\, y\right) + q_i \log\left(y|\, x\right). \tag{13.22}$$

where q_i is the chain rule constant for increasing orderings.

13.3. **Deriving the Bernoulli Law.** We now will apply chain rule (13.19) to the lattice of ordering in Figure 21, where the elements x, y, z are understood to be orderings.

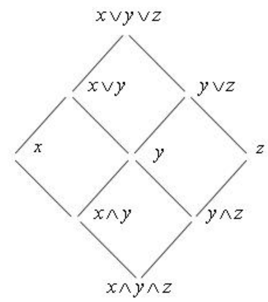

FIGURE 21. A lattice of orderings

We focus on the small diamond in Figure 21, defined by x, $x \vee y$, y, and $x \wedge y$. If we consider the context to be $x \vee y$, then the chain rule (13.19) for this diamond may be written down as:

$$q_d \log\left(x \wedge y \middle| x \vee y\right) = q_d \log\left(x \wedge y \middle| y\right) + q_d \log\left(y \middle| x \vee y\right). \tag{13.23}$$

which reduces to, by way of (13.1) and (13.2),

$$q_d \log\left(x \middle| y\right) = q_d \log\left(x \middle| y\right) + q_d \log\left(y \middle| y\right). \tag{13.24}$$

which implies that

$$q_d \log\left(y \middle| y\right) = 0. \tag{13.25}$$

It follows from (13.25) and the properties of the logarithm that

$$q_d \log\left(x \middle| y\right) = q_d \log \frac{x}{y}. \tag{13.26}$$

We again focus on the small diamond in Figure 21, defined by x, $x \vee y$, y, and $x \wedge y$. If we now consider the context to be $x \wedge y$, then the chain rule (13.22) for this diamond may be written down as:

$$q_i \log\left(x \vee y \middle| x \wedge y\right) = q_i \log\left(x \vee y \middle| y\right) + q_i \log\left(y \middle| x \wedge y\right). \tag{13.27}$$

which, by way of (13.1) and (13.2), reduces to

$$q_i \log\left(y \middle| x\right) = q_i \log\left(y \middle| y\right) + q_i \log\left(y \middle| x\right). \tag{13.28}$$

which implies that

$$q_i \log\left(y \middle| y\right) = 0. \tag{13.29}$$

It follows from (13.29) and the properties of the logarithm that

$$q_i \log\left(y \middle| x\right) = q_i \log \frac{y}{x}. \tag{13.30}$$

Now, we may go, by way of (13.30), from an ordering x to y, and then, by way of (13.26), go from y back to x again. The taking of this path should be consistent, in that the net gain in ordering is zero, which leaves with the equality:

$$q_i \log \frac{y}{x} + q_d \log \frac{x}{y} = 0, \tag{13.31}$$

or, equivalently,

$$q_i \log \frac{y}{x} = -q_d \log \frac{x}{y} = q_d \log \frac{y}{x}. \tag{13.32}$$

It follows from (13.32) that for the lattice of order, consistency demands that the constants q_i and q_d must be equal, which leaves us with the Bernoulli law, which for $x < y$, assigns the valuation

$$q \log \frac{y}{x}. \tag{13.33}$$

for an increase in ordering, and

$$q \log \frac{x}{y}. \tag{13.34}$$

for a decrease in ordering.

If the ordering elements x_1, x_2, etc., are numbers on the positive real, where $x_1 < x_2 <$ etc..., then x_i will tend to ∞, as $i \to \infty$. However, if the ordering elements x_1, x_2, etc., are numbers on the negative real, where $x_1 < x_2 <$ etc..., then x_i will tend to 0, as $i \to \infty$.

This then explains why loss aversion, a phenomenon belonging to the positive Bernoulli law, in which losses are weighted heavier than commensurate gains, in the negative Bernoulli law, changes to the phenomenon of debt relief in which gains are weighted heavier than commensurate losses.

14. Discussion

In this fact sheet we have presented the eight supporting contacts, in chronological order, that led us, finally, to the belief that the Bayesian decision theory, just like the Bayesian probability and information theories, is Bayesian in the strictest sense in the word; that is, an inescapable consequence of the desideratum of consistency. This belief led us to consider that the Bernoulli law, the only remaining degree of freedom in our decision theory, might be more fundamental then we initially had thought.

Because our initial justification for the Bernoulli law had come from the observation that this law, in the guise of the Weber-Fechner and the Steven's power law, had been demonstrated by psycho-physics to be an appropriate model for the way we humans perceive the increments in sensory stimuli, in terms of sensation strength. So, if monetary outcomes are considered to be a sensory stimuli, in the most abstract sense of the word, then it would follow the Bernoulli law would be the most appropriate model for the way we humans perceive the increments in monetary wealth, in terms of their utilities.

The tipping point, where the Bayesian decision theory transitioned from an intuitive idea to something more fundamental, came for us when, having found the skewness intervals, we were re-analyzing the Kahnmenan and Tversky data on the psychological preferences in certainty bets.

It was found that Bayesian decision algorithm confirmed most of the reported preferences. Nonetheless, some preferences were forcefully rejected.

For example, for the certainty bet where we have to choose between a certain gain of 3000, and a probability $p = 0.8$ of gaining 4000 and a probability of $1 - p = 0.2$ of gaining nothing, it is found that 80% of the $N = 95$ subjects preferred the certain outcome, [33].

For an initial wealth of $m = 1000$, it is found that, by way of the Bernoulli law and skewness interval, that the fair probability for this certainty bet is $p = 0.963$. So, only for probabilities larger than this fair probability, will those with a modest income feel inclined to consider the uncertainty choice, which is in correspondence with the observed preference for the certainty choice.

However, for the certainty bet where we have to choose between a certain loss of 3000, and a probability $p = 0.8$ of losing 4000 and a probability of $1 - p = 0.2$ of losing nothing, it is found that 92% of the $N = 95$ subjects preferred the uncertain outcome, [33].

For an initial wealth of $m = 5000$, it is found that, by way of the Bernoulli law and skewness interval, that the fair probability for this certainty bet is $p = 0.747$. So, only for probabilities smaller than this fair probability, will those with a modest income feel inclined to consider the uncertainty choice, which is in strong contradiction with the observed preference for the uncertainty choice.

Now, for an initial wealth that tends to infinity, both fair probabilities will tend to $p = 0.902$, which is in correspondence with both observed preferences. Nonetheless, we felt that human intuition had erred in the latter experiment, as the both the subjects and we ourselves[47], do not have initial wealths that tend to infinity.

In the study of Kahneman and Tversky's work, we had learned to doubt somewhat the infallibility of the experimental method of hypothetical betting choices[48]. As a consequence, we were put in the position that we put more faith in the Bayesian decision theory, than the reported preferences in the hypothetical betting choices.

[47] For we were, in choice, among the 92% who opted for the uncertainty bet.

[48] See Appendix F.

That is, we trusted the Bayesian decision algorithm to teach our intuition, in those instances where the intuitive 'resolution' is lacking to make clear and crisp choices[49].

Especially so, since we, on the one hand, had eliminated the confounding effect of the skewness of the utility probability distributions, by way of the skewness intervals, and, on the other hand, had taken painstaking care to search out those 'unyielding practical realities', as mentioned in the introduction, that would put our foundations to the test. And it had been found that all these practical realities fell nicely in line with the proposed foundations of the Bayesian decision theory.

Moreover, the resulting decision criteria managed to educate our intuitions in unexpected ways; from the adjusted odds ratio, to the way the skewness intervals were handled, to the way what constituted a fair outcome confidence interval, to the way the fair probabilities curve replicated the qualitative S-curve of the probability weighting functions of Kahneman and Tversky.

So, by analogy, Jaynes' reasoning computer of Bayesian probability theory, [28], had become a decision making computer. And this decision making computer, not much unlike a veteran stock broker, knew when to take his losses and not to throw good money after bad.

This then put the burden on us to provide a proof of the fundamentalness of the Bayesian decision theory. Because the history of Bayesian probability has taught us that the usefulness of a theory, in terms of its practical and beautifully intuitive results, in the absence of a compelling axiomatic basis, provides no safeguard against attacks by those who choose to close their eyes to this usefulness[50].

It may be read in Jaynes' [28], that to the best of his knowledge, there are as of yet no formal principles at all for assigning numerical values to loss functions; not even when the criterion is purely economic, because the utility of money remains ill-defined. In the absence of these formal principles, Jaynes final verdict was that decision theory can not be fundamental.

We believe that Jaynes would have approved, would he have been told that his direct descendents, Knuth and Skilling, would be the ones that would provide the Bayesian community with the formal principles with which to assign numerical values to loss functions[51].

---

[49]Just like we have learned, having been Bayesians for the past ten years, to trust the Bayesian probability algorithm to teach our intuition, in those instances where the intuitive resolution is lacking to make clear and crisp plausibility assessments.

[50]Note that this historical fact explains why Bayesians have their axiomatic house in such good order. This process started with the work of Cox, [8], was expanded upon by Jaynes, [28], which was then further refined by the work of Knuth and Skilling, [39]. Moreover, the more general axiomatic framework of the latter has enabled them, amongst other things, [41], to bring some order to the field of quantum theory, by showing why this theory is forced to use a complex arithmetic, [22].

[51]Though we think that the research along the lines of [22] would have pleased him even more. Because the deepest driving motivation behind all of Jaynes' work on statistical theory was not

For that is what Knuth and Skilling have done, by providing the lattice theoretical framework, on which quantifications are derived, by way of associativity symmetries on the underlying lattice algebra, which inherits its meaning from the join and meet of its constituting lattice elements.

The Bernoulli law, initially derived by Bernoulli, by way of common sense first principles, has now been derived by way of a quantification on the lattice of ordering; thus, removing the one remaining degree of freedom of the Bayesian decision theory, and, in the process, demonstrating why it is that Bernoulli's law has proved to be so ubiquitous in the field of psycho-physics. Simply, because consistency demands it.

15. A Post Scriptum by the First Author

Now, if the Bayesian decision theory is indeed an inescapable consequence of the desideratum of consistency, as we think it is, then we would have that the whole human gamut of plausibility and relevancy perception, and the consequent decision making process, when it comes to decisions of a monetary nature, tend to adhere to the primary first principle of consistency. And, not much unlike Fechner, [15], we are suddenly struck by the implied harmony between the physical and mental universes.

For, one might argue that the material world is just a dead mechanism, with life and consciousness occurring only as an incident and only as incidental and fully predetermined by-products of mechanistic laws. In such a soulless world, mental irrationality is not that great of a mystery. For we humans are, obviously, just flawed mechanisms, accidents of an uncaring nature.

The opposite argument, then, would be that Nature itself has a deep preference for consistency[52]. And we humans, being creatures of Nature, feel a deep need to approach, however faltering, Nature's perfect consistency. This need is such, that some of us may even commit themselves to a rigorous and lifelong program of mental training[53]. In this opposite world view, mental irrationality becomes the mystery, as opposed to it being the iron standard.

In the latter view of the world, it becomes apparent why it is that the work of Kahneman and Tversky, in its aggresive and triumphant self-assertion, may inspire

just the desire for more powerful practical methods of inference. It was rather the conviction that progress in basic understanding of physical law, prevented for fifty years by the positivist Copenhagen philosophy, could be resumed only by a drastic modification of the view of the world then taught to physics. Jaynes was of the belief that the mathematics of quantum theory described in part physical law, in part human inference, all scrambled together in such a way that nobody had seen how to seperate them. Jaynes had become convinced that this unscrambling would require that the probability theory itself should be reformulated along the lines of the Bayesian probability theory, [26, 27].

[52] The apple always drops downwards from the tree. Light is always bended by the gravity well of a black hole. Electricity currents always flows from the negative charge to the positive. Etc...

[53] Those who do are typically called scientists.

such ire. Because of its characterization of humans as cognitive cripples, it dismisses our highest mental aspirations, and, by so doing, it triggers our defensive fight or flight reaction.

We may read in [48], that Edwards himself, who was Kahneman and Tversky's mentor, struggled for many years to make his peace with the work of his former pupils, but that he, nonetheless, never succeeded in doing so.

Even Jaynes, veteran of many years of the polemical warfare that was the great schism between Bayesians and frequentists, who, like Newton, could crush his intellectual opponents with a playful swoop of his mighty paw, seems to have been defeated by the impenetrability of the Kahneman and Tversky work. For we may read in [28], which was to be his statistical legacy, that the Kahneman and Tversky psychological experiments are inherently silly, and that if you call something Bayesian, it need not necessarily be so. This type of argumentation is in total contradiction to Jaynes' usual high form, [24, 26], which leads us to the belief that Jaynes decided that he just could not be bothered by the inanity of it all.

We ourselves too, if left to our own devices, would have abandoned the critical analysis of the Kahneman and Tversky work a long time ago, seeing that this work hardly admits any point of attack because of its very amorphousness[54]. Were it not for a very fortunate, though, at times, still painful, serendipity, which, eventually, forced us to formulate a Bayesian answer to the behavioristic economics paradigm.

We say eventually, because we tried our hardest to steer ourselves away from the Kahneman and Tversky work, which never failed to depress us.

Our first line of escape was the development of a new class of C-splines, an explicit base for B-splines, and a non-informative prior with does not needlessly penalize these highly parametrized regression models, thus, solving the problem of

[54]It took these authors two years to formulate an articulate answer to the Kahneman and Tversky charge that humans reason in a fundamentally non-Bayesian manner. And we invite all Bayesians to take a look at the decision theoretical paper [33]. In this paper Kahneman and Tversky define the weighting function π as some monotonic increasing function in p. For the weighted probability $\pi(p)$, it is said that the impossible and certain events are weighted, respectively, such that $\pi(0) = 0$ and $\pi(1) = 1$. The property of subcertainty then is that we may have that

$$\pi(p) + \pi(1-p) < 1.$$

But this implies the possibility of

$$\pi(p) + \pi(1-p) < \pi(1)\,,$$

which then would imply that the subjectively weighted probabilities of two mutually exclusive and exhaustive propositions do not add up to certainty, which is in violation of the sum rule. Moreover it is stated by Kahneman and Tversky that this violation is an 'essential element of people's attitude to uncertain events'. However, when reading [33], do not make the mistake we ourselves initially made, to read it like it actually were a work of statistics. By doing so, one may invite a massive mental disconnect, as one's statistical training will read meaning where there is none, which further down the line, as Kahneman and Tversky take their argument to its inevitable conclusion, will leave the statistically trained reader both confused and bewildered. But rather, read it like an essay on statistics by two under-graduate students, who, unfettered by any actual statistical knowledge, let their fancy take them where ever it may lead.

over-fitting, [12, 13, 44]. However, seeing that we did not have a PhD in numerical analysis, but a PhD in risk perception, this could do little in terms of alleviating our plight.

The second line of escape, which was closer to the mark, for relevancy judgments play an important role in risk communication, was Knuth's Inquiry Calculus, which at the time was still a work in progress, as the specific product rule for the relevancy measure was still lacking. We desperately needed this Bayesian information theory to work, as it offered us a way out of our predicament. So, we tinkered away, until we had a working specific product rule for relevancies, and when Knuth and Skilling came with their [39], we could formulate a formal proof to accompany our tinkering, [14].

Furthermore, the Bayesian information theory also provided us with a non-trivial application of our Inner Nested Sampling algorithm, which is a specific implementation, for Dirichlet distributions, of Skilling's 2004 general Nested Sampling framework[55]. This Inner Nested Sampling algorithm, on which the jury is still out[56], promised to give the information theoretical thesis work the additional technical gravitas, with which we could, in all good conscience, consider our PhD mission accomplished.

But these were all, however gratifying their results, in the final analysis, moves of desperation. For the work of Kahneman and Tversky, was still there, ever triumphant in its drab belligerence; denying us, by way of its anti-Bayesian stance, entry into the decision theoretical field. It was only when we saw our contract renewed, through the European Commission's Seventh Framework Program, that we were finally forced to face our nemesis.

The work package to which we had been assigned called for a behavioristic economical analysis of multi-hazard events. And it was only through the kind, but nonetheless unrelenting, insistence of our work package leader[57] that we came to the

[55]Nested Sampling is a Bayesian Monte Carlo sampling scheme, which represents a quantum leap in the way we numerically evaluate highly variate integrals. Integrals that until now defied all direct evaluation, by way of the curse of dimensionality, may now be evaluated as a matter of course, and Nested Sampling framework, as an added bonus also will provide the user, if needed, with a set of properly weighed random samples, [54].

[56]For we still have to submit, as was promised, to Skilling a revised version in which a more generous amount of graphs is included. So as to better guide the unsuspecting reader through the somewhat involved geometrical arguments, which, on the one hand, take advantage of the n-simplex form of the parameter space of the Dirichlet distribution, and, on the other hand, use the surface of the unit n-sphere to obtain uniformly sampled differentials of the contour of the initial likelihood space.

[57]We here acknowledge our debt of gratitude to Nadejda Komendantova, for doing what a work package leader ought to do. First by encouraging us to stay on point, for information theory is no decision theory, as we ourselves, at the time, were painfully aware. And, later on, when the first draft of the Bayesian decision theory had seen the light of day, through the first three case studies presented in this fact sheet, shielding this theory against the attacks from a somewhat underwhelmed scientific reviewer, who thought it all to be too abstract, with so much still left

Bayesian decision theory, whose structure allowed us to finally begin to formulate a Bayesian answer to the Kahneman and Tversky work. An answer that was long overdue.

Acknowledgments: The research leading to these results has received partial funding from the European Commission's Seventh Framework Program [FP7/2007-2013] under grant agreement no.265138.

References

[1] Aczel, J.: *Lectures on Functional Equations and Their Applications*, Academic Press, New York, (1966).

[2] Allais M.: *L'Extension des Theories de l'Equilibre Economique General et du Rendement Social au Cas du Risque*, Econometrica, 21, 269-290, (1953) .

[3] Allais M.: *Fondements d'une Theorie Positive des Choix Comportant un Risque et Critique des Postulates et Axiomes de l'Ecole Americaine*, Colloque Internationalle du Centre National de la Recherche Scientifique, No. 36, (1952).

[4] Allais M.: *Le Comportement de l'Homme Rationel devant le Risque, Critique des Postulates et Axiomes de l'Ecole Americaine*, Econometrica 21,503-546, (1953).

[5] Allais M.: *An Outline of My Main Contributions to Economic Science*, Nobel Lecture, December 9, (1988).

[6] Bernoulli D.: *Exposition of a New Theory on the Measurement of Risk*. Translated from Latin into English by Dr Louise Sommer from 'Specimen Theoriae Novae de Mensura Sortis', Commentarii Academiae Scientiarum Imperialis Petropolitanas, Tomus V, 175-192, (1738).

[7] Bernstein W.: *A Splendid Exchange; How Trade Shaped the World*, Grove Atlantic Ltd., (2008).

[8] Cox R.T.: *Probability, Frequency and Reasonable Expectation*, American Journal of Physics, 14, 1-13, (1946).

[9] Cox R.T.: *The Algebra of Probable Inference*, John Hopkins University Press, Baltimore MD, (1961).

[10] Edwards W.: *The Theory of Decision Making*, Psychological Bulletin, Vol 51., No. 4, (1954).

[11] Ellsberg D.: *Risk, Ambiguity, and the Savage Axioms*, Quarterly Journal of Economics 75, 643-699, (1961).

[12] Erp van H.R.N., Linger R.O., and Gelder van P.H.A.J.M.: *Constructing Cartesian Splines*. The Open Numerical Methods Journal, 3, 26-30, (2011). But we recommend to search for the unmutilated arXiv version of this article: arXiv:1409.5955 [math.NA], (2014).

[13] Erp van H.R.N., Linger R.O., and Gelder van P.H.A.J.M.: *Deriving Proper Uniform Priors for Regression Coefficients, Part II*, arXiv:1308.1114 [stat.ME], (2013).

[14] Erp van H.R.N.: *Uncovering the Specific Product Rule for the Lattice of Questions*, arXiv:1308.6303 [stat.ME], (2013).

[15] Fancher R.E.: *Pioneers of Psychology*, W. W. Norton and Company, London, (1990).

---

unspecified. As an aside, we made our peace with this reviewer, when at the very end of the project he complimented us, in a final personal encounter, on the obvious mathematical prowess that had gone into the construction of the insurance examples. And looking back, we can only be thankful that this reviewer was as stern as he was. For had he not been so, we might just have left the Bayesian decision theory as was, satisfied with a false sense of completion; thus, denying ourselves the richness that was hidden there in its depths.

[16] Fechner G.J.: *Elemente der Psychophysik, 2 vols.; Vol. 1 translated as Elements of Psychophysics*, Boring, E.G. and Howes, D.H., eds. Holt, Rinehart and Winston, New York, (1966).

[17] Finetti de B.: *Theory of Probability, 2 vols.*, J. Wiley and Sons, Inc., New York, (1974).

[18] Georgescu-Roegen N.: *Utility, Expectations, Measurability, and Prediction*, Paper read Econometric Soc., September (1953).

[19] Go S.C.: *Marine Insurance in the Netherlands 1600-1870: A Comparative Institutional Approach*, Groningen, (2009).

[20] Good, I.J.: *Probability and the Weighing of Evidence*, C. Griffin and Co., London, (1950).

[21] Good, I.J.: *The Contributions of Jeffreys to Bayesian Statistics*, in Zellner, A. ed., Bayesian Analysis in Econometrics and Statistics, North-Holland Pub. Co., Amsterdam, (1980).

[22] Goyal P., Knuth K. H., and Skilling J.: *Origin of Complex Quantum Amplitudes and Feynman's Rules*, Phys Rev. A 81 (2010), 022109, arXiv:0907.090 [quant-ph].

[23] Hall P.: *The Bootstrap and Edgeworth Expansion*, Springer-Verlag (1992).

[24] Jaynes E.T.: *Confidence Intervals vs Bayesian Intervals; Reply to Kempthorne's Comments*, W.L. Harper and C.A. Hooker, eds. Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Reidel Publishing Co., Dordrecht, Holland, (1976).

[25] Jaynes E.T.: *Where Do We Stand on Maximum Entropy?*, in Levine R.D. and Tribus M., eds., The Maximum Entropy Formalism, M.I.T. Press, Cambridge MA, (1978).

[26] Jaynes E.T.: *Some Random Observations*, Synthese 63 115-138, (1985).

[27] Jaynes E.T.: *A Backward Look into the Future*, Jaynes' retirement speech, on-line available.

[28] Jaynes E.T.: *Probability Theory; the Logic of Science*. Cambridge University Press, (2003).

[29] Jeffreys H.: *Theory of Probability Theory*, Clarendon Press, Oxford, (1939).

[30] Hudson M.: *The New Road to Serfdom; An Illustrated Guide to the Coming Real Estate Collapse*, Harper's Magazine, (May 2006).

[31] Kahneman D. and Tversky A.: *Subjective Probability: a Judgment of Representativeness*, Cognitive Psychology 3, 430-454, (1972).

[32] Kahneman D. and Tversky A.: *On the Psychology of Prediction*, Psychological Review, Vol. 80, No. 4, (1973).

[33] Kahneman D. and Tversky A.: *Prospect Theory: an Analysis of Decision Under Risk*, Econometrica, 47(2), 263-291, (1979).

[34] Kahneman D: *Maps of Bounded Rationality: A Perspective on Intuitive Judgement and Choice*, Nobel Lecture, December 8, (2002).

[35] Keynes J.M.: *A Treatise on Probability*, MacMillan, London, (1921); Reprinted by Harper and Row, New York (1962).

[36] Knuth K.H.: *Lattice Duality: the Origin of Probability and Entropy*. Neurocomputing 67;245-274, (2004).

[37] Knuth K.H.: *Information and Entropy, Power Point Presentation*, on-line available, (2008).

[38] Knuth K.H.: *Measuring on Lattices*, (2009).

[39] Knuth K.H. and Skilling J.: *Foundations of Inference*, arXiv: 1008.4831v1 [math.PR], (2010).

[40] Knuth K.H.: *A Derivation of Special Relativity from Causal Sets*, arXiv: 1005.4172v2 [math-ph], 29 Aug. (2010).

[41] Knuth K.H.: *Information-Based Physics: An Observer-Centric Foundation*. Contemporary Physics, (Invited Submission). doi:10.1080/00107514.2013.853426. arXiv:1310.1667 [quant-ph]

[42] Laplace P.S.: *Essai Philosophique sur les Probabilites*, Courcier Imprimeur, Paris, (1819).

[43] Lindgren B.W.: *Statistical Theory*, Chapman & Hall, Inc., New York, (1993).

[44] Linger R.O., Erp van H.R.N., and Gelder van P.H.A.J.M.: *Constructing Explicit B-Spline Bases*, arXiv:1409.3824, (2014).

[45] McGlothlin W.H.: *Stability of Choices among Uncertain Alternatives*, American Journal of Psychology, 69, 604-615, (1956).

[46] Mongin P.: *Expected Utility Theory*, Prepared for the Handbook of Economic Methodology (slightly longer version than the published one), Davis J., Hands W., and Maki U., eds., London, Edward Elgar, (1997).

[47] Neumann von J. and Morgenstern O.: *Theory of Games and Economic Behavior*, Princeton, Princeton Univer. Press, (1944).

[48] Phillips L.D. and von Winterfeldt D.: *Reflections on the Contributions of Ward Edwards to Decision Analysis and Behavioral Research*, Working Paper LSEOR 06.86, Operational Research Group, Department of Management London School of Economics and Political Science, (2006).

[49] Polya G.: *How to Solve It*, Princeton University Press, (1945). Second paperbound edition by Doubleday Anchor Books, (1957).

[50] Polya G.: *Mathematics and Plausible Reasoning, 2 vols: Induction and Analogy in Mathematics*, Patterns of Plausible Inference, Princeton Press, (1954).

[51] MacKay D.J.C.: *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge, (2003).

[52] Masin S.C., Zudini V., and Antonelli M.: *Early Alternative Derivations of Fechner's Law*, Journal of Behavioral Sciences, 45, 56-65, (2009).

[53] Rosenkrantz R.D.: *Inference, Method, and Decision: Towards a Bayesian Philosophy of Science*, D. Reidel Publishing Co., Boston, (1977).

[54] Skilling J.: *Nested Sampling for Bayesian Computations*, Proc. Valencia, ISBA 8th World Meeting on Bayesian Statistics, (2006).

[55] Skilling J.: *The Canvas of Rationality*, Bayesian Inference and Maximum Entropy Methods in Science and Engineering-28^{th} International Workshop, edited by M. de Souza Lauretto, C.A. de Braganca Pereira, and J.M. Stern, (2008).

[56] Slovic P., Finucane M.L., Peters E., and MacGregor D.G.: *Risk as Analysis and Risk as Feelings: Some Thoughts About Affect, Reason, Risk and Rationality*, Risk Analysis 24(2), (2004).

[57] Stevens S.S.: *To Honor Fechner and Repeal His Law*, Science, New Series, Vol. 133, No. 3446, 80-86, (1961).

[58] Tribus M.: *Rational Descriptions, Decisions, and Designs*, Pergamon Press, New York, (1969).

[59] Tversky, A., and Kahneman, D.: *Rational Choice and the Framing of Decisions*. Journal of Business, 59, S251-S278, (1986).

[60] Tversky A. and Kahneman D.: *Advances in Prospect Theory: Cumulative Representation of Uncertainty*, Journal of Risk and Uncertainty, 5: 297-323, (1992).

Appendix A. Variance Preferences

In the Bayesian decision theory bounds of confidence intervals are compared with each other. So, in the Bayesian framework both the expectation values and standard deviations, or, equivalently, variances, of the utility probability distributions are taken into account in the making of decisions. It turns out that this suggestion was also made by both Allais [3, 4, 2] and Georgescu-Roegen[58] [18].

Moreover, Allais constructed his famous paradox for the sole purpose of demonstrating the psychological reality of 'variance preferences', [5]. An Allais paradox may go as follows. Assuming linear utilities for the value of money, we have bets D_1 and D_2, which have the utility probability distributions:

$$P(u|\, D_1) = \begin{cases} 1, & u = 1.000.000 \end{cases} \tag{A.1}$$

and

$$P(u|\, D_2) = \begin{cases} 0.5, & u = 0 \\ 0.5, & u = 4.000.000 \end{cases} \tag{A.2}$$

Based on the utility probability distributions (A.1) and (A.2), people tend to prefer bet D_1 over D_2, even though the utility expectation value under bet D_1 is much smaller than under bet D_2,

$$E(u|\, D_1) = 1.000.000 < 2.000.000 = E(u|\, D_2)\,, \tag{A.3}$$

which is in contradiction with the basic premise of expected utility theory that people will choose that bet which maximizes the expected utility.

Allais contributed this finding to the fact that the variance under bet D_1 is zero, while under bet D_2 it is much greater than zero; what holds for the variances, also holds for standard deviations:

$$\text{std}(u|\, D_1) = 0 << 2.000.000 = \text{std}(u|\, D_2)\,. \tag{A.4}$$

These standard deviations, together with their corresponding means, (A.3), convey that D_1 assuredly will lead to a great gain in utility; whereas under D_2 there is a very real chance of not winning anything at all.

People not only try to maximize the expectation value of utility, they also take into account the variances of the respective utility probability distributions. Hence, the name variance preferences, that is, preferences between decisions based upon the variance, or, equivalently, the standard deviations of the utility probability distributions[59].

Allais' paradox stands prominent among the paradoxes which are used to dismiss excepted utility theory, [46]. This is somewhat ironic, as it is Allais himself who showed us the way out by pointing out that, together with the expected value, the

[58] We were unable to find Georgescu-Roegen's article, referenced in [10], on-line.

[59] Allais states that decisions ought to be taken on the basis of all the information present in the utility probability distribution, $\psi(\gamma)$, by way of some function h, Eq.6, [4]. However, Allais does not proceed to give suggestions as to the form and shape of this function h, at least, as far as we are aware. The lower and upper bounds of the utility probability distributions, as used in this fact sheet, are possible examples of such functions h.

variances and higher order moments of the utility probability distributions should also be taken into account, we quote[60] [5]:

> In the Theory of Games, von Neumann and Morgenstern presented both a method for determining cardinal utility and a rational rule of behavior. Both are based on the consideration of an index which may be called the neo-Bernoullian utility index[61]. The theory devised by von Neumann and Morgenstern demonstrates the existence of this index from a system of postulates, and they identified it with cardinal utility in Jevons' sense. According to them, in order to be rational, any operator must maximize the mathematical expectation of this index.
>
> This stance struck me as being unacceptable because *it amounts to neglecting the probability distribution of psychological values around their mean*, which precisely represents the fundamental psychological element of the theory of risk.
>
> I illustrated my argumentation through counter-examples; one of them became famous as the *'Allais Paradox'*. In fact, the *'Allais Paradox'* is paradoxical in appearance only, and it merely corresponds to a very profound psychological reality, *the preference for security in the neighborhood of certainty.*

The main reason that the concept of variance preferences never caught on is probably because Allais failed to provide an explicit function by which monetary outcomes could be transformed to utilities. Thus, preventing Allais to proceed with the constructing of utility probability distributions and the computation of their variances.

We can only guess as to why Allais disqualified Bernoulli's law as a possible candidate utility function. It may be that Allais disqualified Bernoulli's function because of the latter's oversight to realize the importance of the variances of the utility probability distributions as a criterion of action.

---

[60] Italics are by Allais himself.

[61] Note that the method for determining cardinal utility, mentioned by Allais, refers to the utility measurement scheme which is proposed by von Neumann and Morgenstern in their [47]. This measurement scheme is very much different from the one that was originally proposed by Bernoulli [6]. So much so, that we would opt to designate the resulting utility index to be 'non-Bernoullian', rather than neo-Bernoullian. Bernoulli derives his utility function, or, equivalently, the Weber-Fechner law, by way of three simple considerations. This utility function is then used to compute the expectation of utilities (as opposed to the expectation of monetary outcomes). Von Neumann and Morgenstern, however, postulate, as we believe, a more opague axiomatic system, from which they then derive that the utility indices, which are to be compared, necessarily must take the form of expectation values. A result which is in contradiction with empirical observations, as Ellsberg and Allais have shown with their paradoxes. See [10] for a simple application of the von Neumann and Morgenstern utility assignment scheme.

Or it may be that he thought the problem to be intractable, as also perceived to need to assign subjective probability values to the 'objective' frequentistic probabilities of orthodox statistics.

This then would not only constitute another oversight on the part of Bernoulli, as Bernoulli himself had not perceived this need[62], but it would also compound the problem of assigning moral values to objective monies, seeing that one also would have to assign moral values to objective frequencies.

Appendix B. A Prerequisite Statistical Language

The Bayesian decision theory presented in this thesis is just Bernoulli's expected utility theory, with the intuitive adjustment that we base our decisions on the confidence bounds of our utility probability distributions[63], instead of, as was initially proposed by Bernoulli, their means.

It can hardly be over-stated how amazing it is that Bernoulli got so much of it right in his 1738 essay. Considering that at the time statistics was still in its early infancy.

In 1738 mathematicians where still trying to formulate a solution to the problem of inverse probabilities. The problem of inverse probabilities may be paraphrased as follows, [25]. If, for some probability p, the probability distribution of r successes in n trials is given as

$$p(r|\,n,p) = \frac{n!}{r!\,(n-r)!} p^r\,(1-p)^{n-r}\,.$$

Then what does the observing of r successes in n trials tell us about the probability p, if this probability is unknown?

Bayes was the first one to actually solve the problem of inverse probabilities, as he derived in 1763, for all intents and purposes, the beta distribution:

$$p(p|\,r,n) = \frac{(n+1)!}{r!\,(n-r)!} p^r\,(1-p)^{n-r}\,,$$

However, it was Laplace, in 1774, who took the problem of inverse probabilities to its greatest generality; thus, starting the field of Bayesian statistics.

Laplace posed the following question. If for some set of parameters $\{A\}$, the probability distribution of some other set of parameters $\{B\}$ is given as

$$p(\,\{A\}|\,\{B\})\,.$$

---

[62]This is because Bernoulli regarded probabilities, just like Laplace who came after him, to be a state of knowledge, rather than a limiting frequency of an imaginary infinity of replications of some experiment. And a state of knowledge is always 'subjective', [28].

[63]This adjustment is in the spirit of Allais' suggestion of variance preferences.

Then what does the observing of $\{B\}$ tell us about the probability of $\{A\}$, if $\{A\}$ is unknown? And the solution given by Laplace is the well-known Theorem of Bayes[64]:

$$p(\{B\}|\{A\}) = p(\{B\})\,\frac{p(\{A\}|\{B\})}{p(\{A\})},$$

So, we see that the scientific field of statistics, in 1738, was still very much a field in development.

It was only in 1809 that the normal distribution,

$$p\left(x|\,\mu,\sigma\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right].$$

was discovered by Gauss[65]. This normal law enabled Laplace to finally derive his central limit theorem, which he published in its final form in 1812.

The central limit theorem states that the sum of any x_i observations, for $i = 1,\ldots,n$, will tend to have a normal probability distribution, having a mean of $n\mu$ and a standard deviation of $\sqrt{n}\sigma$, as n tends to infinity, irrespective, of the actual probability law that 'generated' these x_i; where μ and σ are the mean and standard deviation of the generating probability law of the x_i.

As a corollary of the central limit theorem it may be found that if we have just the one observation x, that is, $n = 1$, and if that one observation is generated by the normal law. Then, this sample will have a value somewhere in the range between

$$\left(\mu-\sigma,\mu+\sigma\right),$$

with probability

$$\int_{\mu-\sigma}^{\mu+\sigma}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right]dx = 0.68,$$

which, then, is the rationale behind the ubiquitous use of the sigma confidence intervals.

So, the reason that Bernoulli did not, or better yet, *could not*, take the higher order cumulants of the utility probability distributions into account in his decision theoretical program, was simply because the prerequisite statistical language to think along the lines of confidence bound maximizations was still lacking at the time, when he wrote his 1738 essay.

---

[64] We summarize, the beta distribution is the Bayes' distribution and Bayes' Theorem is Laplace's Theorem.

[65] Gauss himself, at the time, did not realize the importance of the normal law, which he had just derived. But it was Laplace who immediately saw its importance, as he had been studying this normal law himself, through the limiting behavior of the binomial distribution,

$$p\left(r|\,n,p\right) = \frac{n!}{r!\left(n-r\right)!}p^r\left(1-p\right)^{n-r},$$

as n tends to infinity.

But we can imagine, that Bernoulli, had he been here today with us, would have been quite delighted to see that his 1738 proposal, with the proper adjustments, which allow for all the statistical discoveries that have been made from 1738 until now, has stood the test of time so gracefully.

As it will lead, as we will demonstrate in this fact sheet, to very intuitive results, when applied to problems of choice.

Appendix C. Utility Probability Distributions Insurer

Let the insurance company have an initial wealth of M. If the customer pays the insurance premium P and i contingencies occur in conjunction, then the increment in the amount of money for a given outcome O_i is

$$\Delta M_i = P - iL \tag{C.1}$$

Then by way of (4.3) and (C.1), we may construct the following conditional utility distribution, used for mapping outcomes to utilities:

$$p(u|\, O_i, D_1) = \begin{cases} 1, & u = \log \frac{M+P-iL}{M} \\ 0, & u \neq \log \frac{M+P-iL}{M} \end{cases} \tag{C.2}$$

or, equivalently,

$$p(u|\, O_i, P, D_1) = \delta\left(u - \log \frac{M+P-iL}{M}\right), \tag{C.3}$$

where δ is the Dirac delta function:

$$\delta(u-c)\, du = \begin{cases} 1, & u = c \\ 0, & u \neq c \end{cases} \tag{C.4}$$

Because of (C.4), we have that

$$\int \delta(u-c)\, f(u)\, du = f(c)\,. \tag{C.5}$$

This property of the Dirac delta enables us to make a one-on-one mapping, from outcomes to utilities.

By way of the product rule, we then may combine (4.2) and (C.3) in order to obtain the bivariate distribution of the utility u and the outcome O_i:

$$\begin{aligned} p(u, O_i|\, P, D_1) &= p(u|\, O_i, P, D_1)\, p(O_i|\, D_1) \\ &= \delta\left(u - \log \frac{M+P-iL}{M}\right) \binom{n}{i} p^i\, (1-p)^{n-i}\,. \end{aligned} \tag{C.6}$$

Marginalizing over the outcomes O_i, we find the utility probability distribution

$$p(u|\, P, D_1) = \sum_{i=0}^{n} \delta\left(u - \log \frac{M + P - iL}{M}\right) \binom{n}{i} p^i \left(1-p\right)^{n-i}. \tag{C.7}$$

In order to get a more intuitive feel for (C.7) we observe that (C.2) is an one-on-one mapping. So, we may make a change of variable by interchanging the label O_i by its corresponding utility value. This then allows us to write (C.7) in the alternative form:

$$P\left(u = \log \frac{M + P - iL}{M}\right) = \binom{n}{i} p^i \left(1-p\right)^{n-i}. \tag{C.8}$$

Though (C.8) initially may seem a more intuitive notation then (C.7), the Dirac-delta notation of (C.7) is to be preferred, as it more closely reflects the fact that we marginalize over the outcomes O_i in order to obtain univariate probability distribution for the utilities u.

Moreover, after one gets used to the Dirac-delta notation, it will be notation (C.8) which becomes awkward to the eye.

If the insurance company does not sell the insurance, that is, decision D_2, then for each outcome O_i, we have that the initial wealth M of the insurance company remains as is. So,

$$p(u|\, O_i, D_2) = \delta\left(u - \log \frac{M}{M}\right) = \delta(u)\,, \tag{C.9}$$

and we find, by way of the product and generalized sum rule, compare with (C.6) and (C.7):

$$p(u|\, D_2) = \delta(u) \sum_{i=0}^{n} \binom{n}{i} p^i \left(1-p\right)^{n-i} = \delta(u)\,, \tag{C.10}$$

or, equivalently, a probability 1 of neither loss nor gain:

$$P(u = 0) = 1. \tag{C.11}$$

We now have constructed the utility probability distributions under both decisions D_1, insure, and D_2, do not insure; respectively, (C.7) and (C.10).

Appendix D. Non-Linear Preferences

Tversky and Kahneman [60], state that non-linear preferences constitute one of the minimal challenges that must be met by any adequate descriptive theory of choice. We shall explain.

If we have a bet which has the following outcome probability distributions

$$p(O|\, D_1) = \Big\{ 1, \qquad O = 1.000.000 \tag{D.1}$$

and

$$p(O|\,D_2) = \begin{cases} 0.99, & O = 5.000.000 \\ 0.01, & O = 0 \end{cases} \tag{D.2}$$

then people will typically prefer the bet D_1 over D_2.

In contrast, if we have a bet which has the following outcome probability distributions

$$p(O|\,D_1) = \begin{cases} 0.90, & O = 1.000.000 \\ 0.10, & O = 0 \end{cases} \tag{D.3}$$

and

$$p(O|\,D_2) = \begin{cases} 0.89, & O = 5.000.000 \\ 0.11, & O = 0 \end{cases} \tag{D.4}$$

then people will typically prefer the bet D_2 over D_1.

Allais gave this example, in a slightly altered form, to demonstrate that Savage's fifth axiom of independence does not hold, [4]. According to Savage's independence axiom, which we do not endorse, we may add for both (D.1) and (D.2) a probability mass of 0.10 to the proposition $u = 0$, while subtracting that same probability mass for the respective propositions $u = 5.000.000$ and $u = 1.000.000$, leading to (D.3) and (D.4), and still maintain the same problem of choice. But this assumption, as one would hope, is shown to be incorrect by the observed reversal in preferences from bet D_1 over D_2 to bet D_2 over D_1.

Now, according to Kahneman and Tversky the example by Allais not only refutes Savage's axiom of independence, but it also shows that the difference between probabilities of 0.99 and 1.00 has more impact on preferences than the difference between 0.10 and 0.11.

Kahneman and Tversky find this observation to be so full of meaning that they deem it to be a psychological phenomenon in and of itself, and proceed to label it as the 'certainty effect', [33], which later turns into 'non-linear preferences', [60]. But for Bayesians the phenomenon of non-linear preferences is not that special and not that new[66].

[66]The particular U-shape of the non-informative Jeffreys' prior for the parameter θ of the beta distribution,

$$p(\theta) \propto \theta^{-1} (1-\theta)^{-1},$$

is a consequence of the non-linearity of probabilities. If we make a change of variable from θ to the log-odds $\omega = \log[\theta/(1-\theta)]$, then it may be found that the implied non-informative prior of the log-odds ω is uniform,

$$p(\omega) \propto \text{constant}.$$

So, log-odds, having the whole infinity of the x-axis at their disposal, are linear; whereas probabilities, being forced to lie within the heavily constricted interval $[0, 1]$, are not.

While working on the German enigma code, during World War II, Good[67] and Turing introduced the 'deciban' measure which is measured in decibans:

$$\text{deciban}(P) = 10 \log_{10} \frac{P}{1-P}, \tag{D.5}$$

and which gives the plausibility of a proposition being true, relative to it not being true, [21].

For undecidedness, that is, for a fifty-fifty change of some hypothesis A being true, we have

$$P = 1 - P = 0.5. \tag{D.6}$$

Substituting (D.6) in (D.5), we find that undecidedness, (D.6), corresponds with

$$\text{deciban}(0.5) = 10 \log_{10}(1) = 0. \tag{D.7}$$

Just as 1 db sound represents the just noticeable difference relative to silence, so a ± 1 deciban change in probability represents the just noticeable difference relative to undecidedness, [20].

Using (D.5), we find that the decibans associated with the probabilities 0.99 and 1.00 are, respectively,

$$\text{deciban}(0.99) = 10 \log_{10} \frac{0.99}{0.01} = 19.96 \tag{D.8}$$

and

$$\text{deciban}(1.00) = 10 \log_{10} \frac{1.00}{0.00} \to \infty. \tag{D.9}$$

Now, (D.9) tells us that a probability 1.00 is a limit case of absolute certainty. Whereas a probability of 0.99 is not, representing just under 20 decibans of evidence for proposition A being true.

So, the difference in evidence for proposition A being true for the probabilities 0.99 and 1.00 is much more than 1 deciban:

$$\text{deciban}(1.00) - \text{deciban}(0.99) = \infty >> 1. \tag{D.10}$$

It may be checked that, (D.5), the probabilities of 0.10 and 0.11 correspond with a less than 1 deciban difference in evidence:

$$\text{deciban}(0.11) - \text{deciban}(0.10) = 0.46 < 1. \tag{D.11}$$

[67]Good wrote about 2000 articles on Bayesian statistics, found throughout the statistical and philosophical literature starting in 1940. Workers in the field generally granted that every idea in modern statistics can be found expressed by him in one or more of these articles; but their sheer number made it impossible to find or cite them, and most are only one or two pages long, dashed off in an hour and never developed further. So, for many years, whatever one did in Bayesian statistics, one just conceded priority to Jack Good by default, without attempting the literature search, which would have required days. Finally, in 1983, a bibliography was provided of most of the first 1517 of these articles with a long index, so it is now possible to give proper acknowledgments of his works up to 1983, [28].

So, if we find that subjects prefer the second bet in the second collection of bets, equations (D.3) and (D.4). Then this also may be interpreted as follows. Subjects are indifferent to the difference in probabilities 0.10 and 0.11, as this difference represents a change of less then 1 deciban in the plausibility of hypothesis A being true, (D.11). So, all things being equal, subjects choose the bet with the highest potential payout of 5.000.000 dollars.

In closing, the deciban, (D.5), represents the intuitive scale on which the plausibility of proposition A being true, relative to it not being true, is judged; that is, the deciban is the scale of the numerically coded plausibilities. Whereas the probability,

$$P = \frac{10^{\frac{\text{deciban}(P)}{10}}}{1 + 10^{\frac{\text{deciban}(P)}{10}}}, \tag{D.12}$$

represents the non-intuitive 'technical' scale, which follows from the quantification of our common sense, [39, 55].

So, if the difference between probabilities of 0.99 and 1.00 has more impact on preferences than the difference between 0.10 and 0.11, as is found in psychological experiments, then this is reflective of the fact that the qualitative properties of the intuitive deciban-scale, up to a certain point, are retained in the transformation to the more technical probability-scale.

We say up to a certain point, because probability theory, which makes use of the technical probabilities, is common sense amplified, having a much higher probability resolution than our human brains can ever hope to achieve. More concretely, we expect that for human probability perception the range of meaningful decibans is bounded somewhere around, say, ± 40 deciban, [28].

Appendix E. Bernoulli's Error

Kahneman dedicates in his Nobel lecture a whole section on 'Bernoulli's error' and on how prospect theory may remedy this error, [34].

In Kahneman's Nobel lecture we may read the following on Bernoulli's error:

> The idea that decision makers evaluate outcomes by the utility of final asset positions has been retained in economic analyses for almost 300 years. This is rather remarkable, because the idea is easily shown to be wrong; I call it Bernoulli's error. Bernoulli's model is flawed because it is reference-independent: it assumes that the value that is assigned to a given state of wealth does not vary with the decision maker's initial state of wealth.

So, Bernoulli's model is claimed to be in error in that it would evaluate outcomes by the utility of final asset positions alone, without taking into account the initial wealth of the decision maker.

But it may be checked, Paragraph 10 of [6], that Bernoulli gives an utility function of the form:

$$u\left(S|\, S_0\right) = q \log \frac{S}{S_0}. \tag{E.1}$$

where, adopting the Kahneman's terminology, S and S_0 are, respectively, the final and initial asset states. Let the asset increment ΔS be defined as

$$\Delta S = S - S_0. \tag{E.2}$$

Then, by substituting (E.2) into (E.1), we obtain the equivalent utility function, (4.3):

$$u\left(\Delta S|\, S_0\right) = q \log \frac{S_0 + \Delta S}{S_0}. \tag{E.3}$$

It follows that in Bernoulli's expected utility theory asset increments ΔS, be they positive or negative, are evaluated relative to the initial wealth S_0 of the decision maker.

This then begs the question: What was it, that led Kahneman to the misguided belief that in Bernoulli's model the gains and losses are not the carriers of utility?

Kahneman and Tversky state the first two tenets[68] of expected utility theory to be, respectively, the tenets of *expectation* and *asset integration*, [33]. The tenet of expectation is

$$U(x_1, p_1; \ldots; x_n, p_n) = p_1 u(x_1) + \cdots + p_n u(x_n)\,. \tag{E.4}$$

The tenet of asset integration states that if w is our current asset position, that is, our initial wealth, then we will accept an uncertain prospect having outcomes x_i if

$$U(w + x_1, p_1; \ldots; w + x_n, p_n) > u(w)\,. \tag{E.5}$$

By substituting (E.4) into (E.5), we obtain the implied asset integration tenet:

$$p_1 u(w + x_1) + \cdots + p_n u(w + x_n) > u(w)\,. \tag{E.6}$$

But then we have that the tenets of expectation and asset integration, as stated by Kahneman and Tversky, are incompatible with Bernoulli's expected utility theory.

For the expectation tenet (E.4), Bernoulli's expected utility theory [6] implies the function u:

$$u(x) = q \log \frac{w + x}{w}, \tag{E.7}$$

whereas, under the asset integration tenet (E.6), the implied function u would be

$$u(x) = q \log x. \tag{E.8}$$

[68]The third, and last, tenet is that of *loss aversion*, which states that u must be some concave down function. The Bernoulli law is concave down.

By way of (E.7), (E.8), and the fact that

$$q \log \frac{w+x}{w} \neq q \log x,$$

it is then demonstrated that the two tenets, as proposed by Kahneman and Tversky in their [33], are incompatible with Bernoulli's expected utility theory.

By dropping the expectation tenet (E.4), while retaining the asset integration in its form of (E.6), we may, very easily, do away with the Kahneman and Tversky inconsistency.

By substituting the implied (E.8) into the asset integration tenet (E.6), we find[69]:

$$p_1 \left[q \log(w + x_1)\right] + \cdots + p_n \left[q \log(w + x_n)\right] > q \log(w), \tag{E.9}$$

Then, by way of the properties of the log function, we may rewrite (E.9) into the equivalent

$$p_1 \left[q \log \frac{w + x_1}{w}\right] + \cdots + p_n \left[q \log \frac{w + x_n}{w}\right] > 0, \tag{E.10}$$

which for any psychologist should be recognizable as the weighted sum of Weber-Fechner utilities[70].

Especially, if those psychologists, like Kahneman and Tversky, explicitly state that the facts of perceptual adaptation were in their minds when they began their joint research on decision making under risk, [34].

So, if it is claimed by Kahneman and Tversky [33] that Bernoulli's model is in error, as it would evaluate outcomes by the utility of final asset states alone, rather than gains or losses. Then we may infer that Kahneman and Tversky did not fully realize[71] that, according to Bernoulli [6], their abstract (E.6) necessarily implies the concrete (E.10).

At the end of the quote on Bernoulli's error[72] we may find the following cryptic footnote by Kahneman:

> What varies with wealth in Bernoulli's theory is the response to a given *change* of wealth. This variation is represented by the

[69]Note that Laplace [42] discusses Bernoulli's suggestion by way of the equivalent inequality:

$$p_1 \log(w + x_1) + \cdots + p_n \log(w + x_n) > \log(w).$$

[70]The Weber-Fechner law is used, amongst other things, to determine the decibel scale of human sound perception, where the Weber constant has been experimentally determined as $q = \frac{10}{\log 10}$.

[71]In all fairness, we ourselves initially thought that we had improved on the insurance example given by Jaynes in his [28], by using the Weber-Fachner law, which we still remembered from our psychology days, rather than Laplace's

$$p_1 \left[\log(w + x_1)\right] + \cdots + p_n \left[\log(w + x_n)\right] > \log(w).$$

But it was professor Han Vrijling, to whom we owe a debt of gratitude, who first pointed us to Bernoulli's [6], and the equivalence of the Weber-Fechner law and the Bernoulli law. It was only then, that we realized that Laplace's formulation is equivalent to (E.9).

[72]Given at the beginning of the chapter.

> curvature of the utility function for wealth. Such a function cannot be drawn if the utility of wealth is reference-dependent, because utility then depends not only on current wealth but also on the reference level of wealth.

We now will offer up an interpretation of what is stated in this footnote; as it may shed some further light on the Kahneman and Tversky position.

Let us assume that Kahneman and Tversky had at least some sense of what is written in, and here we quote Kahneman, [34], 'the brilliant essay that introduced the first version of expected utility theory (Bernoulli, 1738)'; that is, we assume that they were aware of the fact that Bernoulli proposes to use the log function, in some shape or form, in order to assign utilities to outcomes.

Then it may well have been that Kahneman and Tversky were under the wrongful impression that Benoulli's utility function is given as

$$u = q \log \left(w + x \right), \tag{E.11}$$

The first two sentences in the footnote then may be interpreted as expressing the idea that, with differing levels of wealth w, the supposed utility function (E.11) will be more or less linear in a given change of wealth x.

If in the third sentence we let 'current wealth' stand for change in wealth and 'reference level of wealth' for initial state of wealth, then we may interpret it as stating that (E.11) misses the necessary structure to take into account the initial state of wealth in its utility assignments.

Bernoulli's supposed error then would be that he proposed as his utility function (E.11). But Bernoulli proposed (E.7) instead of (E.11).

The erroneous utility function (E.11) is problematic in that it cannot assign a value of zero to a change of wealth of $x = 0$. In Kahneman and Tversky's prospect theory we have that changes in wealth x are assigned values by way of the value function v, where[73]

$$v(x) = 0, \qquad \text{for } x = 0. \tag{E.12}$$

So, if we read in [34]:

> Preferences appeared to be determined by attitudes to gains and losses, defined relative to a reference point, but Bernoulli's theory and its successors did not incorporate a reference point. We therefore proposed an alternative theory of risk, in which the carriers of utility are gains and losses - changes of wealth rather than states of wealth.

[73]Note, Kahneman and Tversky's two-part value function is given as, [60]:

$$v(x) = \begin{cases} x^{\alpha}, & x \geq 0 \\ -\lambda \left(-x \right)^{\beta}, & x < 0 \end{cases}$$

.

> Prospect theory (Kahneman & Tversky, 1979) embraces the idea that preferences are reference-dependent, and includes the extra parameter that is required by this assumption.

Then we may interpret is as saying that the erroneous u cannot assign zero utilities to zero outcomes, whereas Kahneman and Tversky's value function v can.

It would then follow that that which is embraced by prospect theory is the *constraint*[74] (E.12) on the value function v. But this constraint also holds, trivially, for Bernoulli's utility function (E.7).

Now, after having established some tentative understanding into the reasoning process that might have led Kahneman and Tversky to their misunderstanding of Bernoulli's position, and after having provided a possible interpretation of Kahneman's footnote, we may start to wonder: We know how the initial wealth w factors into Bernoulli's expected utility theory, but how does this initial wealth factor into prospect theory?

We quote Kahneman and Tversky [33]:

> The emphasis on changes as the carriers of value should not be taken to imply that the value of a particular change is independent of initial position. Strictly speaking, value should be treated as a function in two arguments: the asset position that serves as a reference point, and the magnitude of the change (positive or negative) from that reference point.

And we could not agree more, though we ourselves would have dropped the 'strictly speaking' modifier, as it weakens the imperative. Kahneman and Tversky continue:

> However, the preference order of prospects is not greatly altered by small or even moderate variations in asset position. . . . Consequently, the representation of value as a function in one argument generally provides a satisfactory approximation.

So, it is Kahneman himself rather than Bernoulli, who does not take explicitly into account the decision maker's initial state of wealth[75].

Appendix F. Bayesian Probability Theory vs. Kahneman and Tversky

The whole of Bayesian probability theory flows forth from two simple rules. The product rule,

$$P(A)\,P(B|\,A) = P(AB) = P(B)\,P(\,A|\,B)\,, \tag{F.1}$$

and the sum rule

$$P(\overline{A}) = 1 - P(A)\,. \tag{F.2}$$

[74]Constraints are not the same as parameters.

[75]As may be checked in the previous footnote.

By way of the product and the sum rule, we may derive the generalized sum rule,

$$P(A + B) = P(A) + P(B) - P(AB)\,. \tag{F.3}$$

If we have that the propositions are exhaustive and mutually exclusive, that is, $B = \overline{A}$, we have that, by way of (F.2) and(F.3),

$$P\left(A + \overline{A}\right) = P(A) + P\left(\overline{A}\right) = 1, \tag{F.4}$$

we may obtain a probability distribution. This probability distribution then may be further generalized to the bivariate probability distribution:

$$\sum_i \sum_j P(A_i B_j) = 1, \tag{F.5}$$

and higher variate probability distributions, which allows us 'marginalize' over any parameter, say, B_j, which is of no direct interest:

$$P(A_i) = \sum_j P(A_i B_j)\,. \tag{F.6}$$

Now, to a non-Bayesian it may seem to be somewhat surprising, that the whole of Bayesian probability theory flows forth from the product and rules. But the whole of Boolean logic, on an operational level, is also captured by the AND- and NOT-operations. These operations correspond, respectively, with (F.1) and (F.2); as these operators combine, with the negation of a NAND-operation, in the OR-operation, which corresponds with (F.3).

Moreover, it may be shown that Boolean logic is just a special limit case of the more general Bayesian probability theory. The operators of Boolean logic combine in a like manner as the operators in Bayesian probability theory. But in Boolean logic propositions can have only the truth values true or false. Whereas in Bayesian probability theory propositions can have plausibility values in the interval $[0, 1]$, where 0 and 1, respectively, correspond with false and true.

So Boolean logic is the language of deduction. Whereas Bayesian probability theory is the language of both induction and deduction; the latter being a limit case of the former, in which we have absolute knowledge about the propositions in play.

We will present in this Appendix first the case of Bayesian probability theory being common sense quantified. We will do this by way of a simple demonstration of the reasoning power of the product and sum rules, (F.1) and (F.2).

We then examine one of the Kahneman and Tversky experiments, which is said to demonstrate that people do not reason as a Bayesian would, that is, consistently[76].

[76]As consistency is the very axiom by which we may derive the product and sum rules of Bayesian probability theory, [8, 28, 39, 55].

F.1. **The case for Bayesian probability theory.** Bayesian statistics is not only said to be common sense quantified, but also common sense amplified[77], having a much higher 'probability resolution' than our human brains can ever hope to achieve [28].

This statement is in accordance with the Kahneman and Tversky finding that, if presented with some chance of a success, p, subjects fail to draw the appropriate binomial probability distribution of the number of successes, r, in n draws. Even though subjects manage to find the expected number of successes, they fail to accurately determine the probability spread of the r successes. Kahneman and Tversky see this as evidence that humans are fundamentally non-Bayesian in the way they do their inference, [31].

We instead propose that human common sense is not hard-wired for problems involving sampling distributions. Otherwise there would be no need for such a thing as data-analysis, as we only would have to take a quick look at our sufficient statistics after which we then would draw the probability distributions of interest.

However, humans do seem to be hard-wired for more 'Darwinian' problems of inference. For example, if we are told that our burglary alarm has gone off, after which we are also told that a small earthquake has occurred in the vicinity of our house around the time that the alarm went off. Then common sense would suggest that the additional information concerning the occurrence of a small earthquake will somehow modify our probability assessment of there actually being a burglar in the house.

We may use Bayesian probability theory to examine how the knowledge of a small earthquake having occurred translates to our state of knowledge regarding the plausibility of a burglary. The narrative we will formally analyze is taken from [51]:

> Fred lives in Los Angeles and commutes 60 miles to work. Whilst at work, he receives a phone-call from his neighbor saying that Fred's burglar alarm is ringing. While driving home to investigate, Fred hears on the radio that there was a small earthquake that day near his home.

[77]If Bayesian inference were not common sense amplified, then it could not ever hope to enjoy the successes it currently enjoys in the various fields of science; astronomy, astrophysics, chemistry, image recognition, etc...

Let

$$B = \text{Burglary}$$
$$\overline{B} = \text{No burglary}$$

$$A = \text{Alarm}$$
$$\overline{A} = \text{No alarm}$$

$$E = \text{Small earthquake}$$
$$\overline{E} = \text{No earthquake}$$

We assume that the neighbor would never phone if the alarm is not ringing and that the radio report is trustworthy too; thus, we know for a fact that the alarm is ringing and that a small earthquake has occurred near the home. Furthermore, we assume that the occurrence of an earthquake and a burglary are independent. We also assume that a burglary alarm is almost certainly triggered by either a burglary or a small earthquake or both, that is,

$$P\left(A|\,B\overline{E}\right) = P\left(A|\,\overline{B}E\right) = P(A|\,BE) \to 1, \tag{F.7}$$

whereas alarms in the absence of both a burglary and a small earthquake are extremely rare, that is,

$$P\left(A|\,\overline{B}\,\overline{E}\right) \to 0. \tag{F.8}$$

Let

$$P(E) = e, \qquad P(B) = b. \tag{F.9}$$

Then we have, by way of the sum rule (F.2),

$$P\left(\overline{E}\right) = 1 - e, \qquad P\left(\overline{B}\right) = 1 - b. \tag{F.10}$$

It follows, by way of the product rule (F.1), as well as (F.7) through (F.10), that

$$\begin{aligned}
P\left(AB\overline{E}\right) &= P\left(A|\,B\overline{E}\right) P(B)\, P\left(\overline{E}\right) \to b\,(1-e)\,, \\
P\left(A\overline{B}E\right) &= P\left(A|\,\overline{B}E\right) P\left(\overline{B}\right) P(E) \to (1-b)\,e, \\
P(ABE) &= P(A|\,BE)\, P(B)\, P(E) \to be, \\
P\left(A\overline{B}\,\overline{E}\right) &= P\left(A|\,\overline{B}\,\overline{E}\right) P\left(\overline{B}\right) P\left(\overline{E}\right) \to 0.
\end{aligned} \tag{F.11}$$

By way of 'marginalization', that is, an application of the generalized sum rule, (F.3), we obtain the probabilities

$$\begin{aligned} P\left(A\overline{B}\right) &= P\left(A\overline{B}\,\overline{E}\right) + P\left(A\overline{B}E\right) \to (1-b)\,e, \\ P(AB) &= P(ABE) + P\left(AB\overline{E}\right) \to b, \end{aligned} \tag{F.12}$$

and

$$P(A) = P(AB) + P\left(A\overline{B}\right) \to b + e - be, \tag{F.13}$$

and

$$\begin{aligned} P\left(A\overline{E}\right) &= P\left(AB\overline{E}\right) + P\left(A\overline{B}\,\overline{E}\right) \to b\,(1-e)\,, \\ P(AE) &= P(ABE) + P\left(A\overline{B}E\right) \to e. \end{aligned} \tag{F.14}$$

The moment Fred hears that his burglary alarm is going off, then there are two possibilities.

One possibility is that Fred may be new to Los Angeles and, consequently, overlook the possibility of a small earthquake triggering his burglary alarm, which will make his prior probability of his alarm going off, (F.13), go to

$$P(A) \to b + e - be = b, \qquad \text{since } e = 0, \tag{F.15}$$

seeing that that the possibility of an earthquake is overlooked.

Fred then assesses, by way of the product rule (F.1),(F.12) and (F.15), the likelihood of a burglary to be

$$P(\,B|\,A) = \frac{P(AB)}{P(AE)} \to \frac{b}{b} = 1, \tag{F.16}$$

which leaves him greatly distressed, as he drives to his home to investigate.

Another possibility is that Fred is a veteran Los Angeleno and, as a consequence, instantly will take into account the hypothesis of a small tremor occurring near his house.

Having an optimistic disposition, Fred does not assign to much weight to the possibility of an earthquake and a burglary occurring both at the same time. So, he updates his prior probability of his alarm going off from (F.13) to

$$P(A) \to b + e - be \to b + e, \qquad \text{since } be \to 0. \tag{F.17}$$

Fred then assesses, by way of the product rule (F.1),(F.12) and (F.17), the likelihood of a burglary to be

$$P(\,B|\,A) = \frac{P(AB)}{P(AE)} \to \frac{b}{b+e}. \tag{F.18}$$

If earthquakes are somewhat more common than burglaries, then Fred, based on his (F.16), may still hope for the best, as he drives home to investigate.

Either way, the moment that Fred hears on the radio that a small earthquake has occurred near his house, around the time when the burglary alarm went off, then, by way of the product rule (F.1) and (F.11) and (F.14), Fred updates the likelihood of a burglary to be

$$P(B|\,AE) = \frac{P(ABE)}{P(AE)} \rightarrow \frac{be}{e} = b. \tag{F.19}$$

In the presence of an alternative explanation for the triggering of the burglary alarm, that is, a small earthquake occurring, the burglary alarm has lost its predictive power over the prior probability of a burglary, that is, (F.9) and (F.19),

$$P(B|\,AE) \rightarrow P(B)\,. \tag{F.20}$$

Consequently, Fred's fear for a burglary, as he rides home, after having heard that a small earthquake did occur, will only be dependent upon his assessment of the general likelihood of a burglary occurring. If we assume that Fred lives in a nice neighborhood, rather than some crime-ridden ghetto, then we can imagine that Fred will be, if not greatly, then at least somewhat, relieved.

One of the arguments made against Bayesian probability theory as a normative model for human rationality is that people are generally numerical illiterate. Hence, the Bayesian model is deemed to be too numerical a model for human inference, [56].

However, note that the Bayesian analysis given here was purely qualitative, in that no actual numerical values were given to our probabilities, apart from (F.7) and (F.8), which are limit cases of certainty and, hence, in a sense, may also be considered to be qualitative.

Moreover, the result of this qualitative analysis seems to be intuitive enough. Indeed, the qualitative correspondence of the product and sum rules with common sense has been noted and demonstrated time and again by many distinguished scientists, including Laplace [42], Keynes [35], Jeffreys [29], Polya [49, 50], Cox [9], Tribus [58], de Finetti [17], Rosenkrantz [53], and Jaynes [28].

F.2. **The case against Bayesian probability theory.** The psychological paradigm of heuristics and biases originated as a reaction to the shortcomings of the mathematical expected utility theory.

In the 1950's it was found that expected utility theory, the then dominant decision theory, failed to adequately model human decision making in certain instances, leading to such paradoxes as the Ellsberg and Allais paradox[78]. Consequently, Edwards

[78]See Appendix A.

and his research team of PhD-students and post-docs took it upon themselves to remedy the situation, [48].

Kahneman and Tversky, both post-docs under Edwards, proposed to construct a systematic theory about the psychology of uncertainty and judgment. In this psychological theory a handful of heuristics would replace the mathematical laws of chance as a more realistic model for subjective judgment of uncertainty. But what started as a parsimonious theory of human inference, consisting of only three heuristics and their associated biases, [34], has proliferated into 20 heuristics and an impressive 170+ biases[79].

Heuristics are said to be mental short cuts or 'rules of thumb' humans use to do inference. It is theorized that, as we do not always have the time or resources to compare all the information at hand, we use heuristics to do inference quickly and efficiently.

However, or so we are warned, even though these mental short cuts will be helpful most of the times, if used carelessly heuristics may lead to heuristic-induced biases, that is, systematic errors in reasoning, [34].

For example, when we use the representativeness heuristic then we estimate the likelihood of an event by comparing it to an existing prototype that already exists in our minds, [32].

Our prototype is what we think is the most relevant or typical example of a particular event or object. The bias associated with the representativeness heuristic is that when making judgments based on representativeness we are likely to overestimate the likelihood that the representative event will occur; just because an event or object is representative does not mean that it is more likely to occur.

In order to demonstrate this bias Kahneman and Tversky performed the following experiment, [32]. They divided the participants in their study up into three groups.

The *base-rate group* was asked to guess the percentage of all first-year graduate students in the following nine fields of specialization: business administration, computer science, engineering, humanities and education, law, library science, medicine, physics, and social sciences.

The base-rate group estimated the highest percentage of graduate students, with 20%, to be in humanities and education, and the second lowest percentage, with 7%, to be in computer sciences.

The *similarity group* was presented with the following profile:

> Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather

[79]Source Wikipedia, search 'heuristics' and 'list of cognitive biases'.

> dull and mechanical, occasionally enlivened by somewhat corny puns and by flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to feel little sympathy for other people and does not enjoy interacting with others. Self-centered, he nonetheless has a deep moral sense.

After which they were asked to rate how similar Tom was perceived to be to the typical graduate student in each of the nine graduate specializations.

The similarity group assigned computer science the highest ranking position, with a mean similarity of 2.1, whereas humanities and education was assigned the second lowest ranking position, with a mean similarity ranking of 7.2.

The *prediction group* was given the same personality sketch of Tom as the similarity group, with the following additional information:

> The preceding personality sketch of Tom W. was written during Tom's senior year in high school by a psychologist, on the basis of projective tests. Tom W. is currently a graduate student.

Then they were asked to rank the nine fields of graduate specialization in order of the likelihood that Tom was now a graduate student in each of these fields.

It was found that 95% of the prediction group judged that Tom was more likely to study computer science than humanities and education.

Since the likelihood rankings of the prediction group closely follow the similarity rankings of the similarity group, whereas they do not follow the base-rate estimates of the base-rate group, Kahneman and Tversky conclude that the representativeness heuristic must have been used by the participants in the prediction group, [32].

However, the use of representativeness does not necessarily imply the representativeness heuristic. We quote Kahneman and Tversky on the representativeness heuristic, [31]:

> Our thesis is that, in many situations, an event A_1 is judged more probable than an event A_2 whenever A_1 appears more representative than A_2. In other words, the ordering of events by their subjective probabilities coincides with their ordering by representativeness.

We now will give a formal analysis of the representativeness heuristic, as proposed by Kahneman and Tversky.

Let

$$\begin{aligned} A_1 &= \text{Computer Science Student} \\ A_2 &= \text{Humanities and Education Student} \\ B &= \text{Profile} \end{aligned}$$

If Tom's psychological profile, B, is more representative of computer science students, A_1, than of humanities students, A_2, then

$$P(B|A_1) > P(B|A_2)\,, \tag{F.21}$$

or, equivalently,

$$\frac{P(B|A_1)}{P(B|A_2)} > 1. \tag{F.22}$$

So, the representativeness heuristic is build upon the thesis:

$$\frac{P(B|A_1)}{P(B|A_2)} > 1 \implies \frac{P(A_1|B)}{P(A_2|B)} > 1, \tag{F.23}$$

where '$\implies$' is the symbol for logical implication.

However, thesis (F.23) is unfounded in that it does not follow directly from the rules of plausible reasoning. Moreover, Kahneman and Tversky seem to intuit as much; seeing that they warn us for the bias of base rate neglect, when using their representativeness thesis, [32].

By taking the conclusion part in the thesis (F.23) as the starting point of a formal Bayesian analysis, we may find, by way of the product rule (F.1), that

$$\begin{aligned} \frac{P(A_1|B)}{P(A_2|B)} &= \frac{P(B)}{P(B)}\frac{P(A_1|B)}{P(A_2|B)} \\ &= \frac{P(A_1B)}{P(A_2B)} \qquad \text{(F.24)} \\ &= \frac{P(A_1)}{P(A_2)}\frac{P(B|A_1)}{P(B|A_2)}. \end{aligned}$$

It follows that the correct thesis, which actually does take into account the base rate, would be

$$\frac{P(B|A_1)}{P(B|A_2)} > \frac{P(A_2)}{P(A_1)} \implies \frac{P(A_1|B)}{P(A_2|B)} > 1. \tag{F.25}$$

Kahneman and Tversky make in [32] the implicit assumption that the reported use of representativeness, that is, an evaluation and use of the odds (F.22), as reported by the participants of the prediction group, necessarily implies their thesis (F.23). This leads them to conclude that the participants must have used their representativeness heuristic.

However, it can be seen that this assumption is incorrect, as the Bayesian thesis (F.25) also makes use of the odds in (F.22) and, thus, representativeness. Moreover, based on the reported use of representativeness and the experimental data, one could make the case that the participants in the experiment intuitively made use of the correct (F.25), instead of the erroneous (F.23).

Kahneman and Tversky report that the prior odds for humanities and education against computer science were estimated by the participants to be, [32]:

$$\frac{P(A_2)}{P(A_1)} \approx 3. \tag{F.26}$$

Now, (F.24), or, equivalently, thesis (F.25), tells us that (F.26) together with

$$\frac{P(B|\, A_1)}{P(B|\, A_2)} > 3 \tag{F.27}$$

implies the conclusion

$$\frac{P(A_1|\, B)}{P(A_2|\, B)} > 1, \tag{F.28}$$

or, equivalently,

$$P(A_1|\, B) > P(A_2|\, B)\,. \tag{F.29}$$

So, if 95% of the participants in the third group judged that Tom was more likely to study computer science than humanities, then we may infer that 95% of the participants deemed the odds-inequality (F.27) to hold, which in our opinion is not that far-fetched[80].

Kahneman and Tversky, however, are of the opinion that the plausibility judgments of the participants in the third group 'drastically violate the rules of the normative [that is, Bayesian] rules of prediction', seeing that the following considerations were ignored by the participants in the prediction group, [32]:

> First given the notorious invalidity of projective personality tests, it is very likely that Tom W. was never in fact as compulsive and as aloof as his description suggests. Second, even if the description was valid when Tom W. was in high school, it may not longer be valid now that he is in graduate school. Finally, even if the description is still valid, there are probably more people who fit that description among students of humanities and education than among students of computer science, simply because there are so many more students in the former than in the latter field.

As to the last consideration, it follows from the Bayesian 'heuristic' (F.25) that the inference of Tom being a computer science student implies the corollary inference that among all the graduate students who fit Tom's description there will be more students of computer science than students of humanities and education. Even if

[80]Indeed, the ranking of computer sciences was, with a mean similarity of 2.1, more than three times higher than the ranking of humanities, which had a mean similarity ranking of only 7.2, [32]. Even though rankings do not translate easily to probabilities, as a *sine qua non*, this ordering of similarity rankings still constitutes corroborating evidence for inequality (F.27) to have held for the participants in the Kahneman and Tversky experiment.

there are many more students in the field of humanities and education than in the field of computer science.

The odds (F.26) represent the ratio of humanities and education students to computer science students. Whereas the odds (F.27) represent the ratio of computer students having a profile like Tom's to humanities and eduction students having a like profile. If the latter odds dominate the former, then we must conclude, by way of (F.24), that there are probably more people who fit Tom's description among students of computer science than among students of humanities and education; that is, $P(A_1B) > P(A_2B)$. Which invalidates Kahneman's and Tversky's last consideration.

Moreover, we have also grown to distrust Kahneman and Tversky's competency somewhat, when it comes to matters of the normative rules of prediction.

This, then, leaves us with the following arguments for the thesis that people tend to neglect the base rate, when taking the mental shortcut (F.23), thus, violating the rules of the normative [that is, Bayesian] rules of prediction:

(1) given the notorious [that is, clinical] invalidity of projective personality tests, it is very likely that Tom W. was never in fact as compulsive and as aloof as his description suggests.

(2) even if the description was valid when Tom W. was in high school, it may not longer be valid now that he is in graduate school.

Now, it would seem that these arguments pertain to some other thesis, namely, that the participants of the prediction group should have disregarded the description of Tom altogether, as no data was actually given. But this alternative thesis, apart from it not being the issue, is not all that compelling.

Because, if, in answer to the first argument, we filter out those qualifications which might point to compulsiveness and aloofness, then we are left with the following profile for Tom:

- high intelligence,
- dull and mechanical writing,
- lacking in true creativity,
- corny puns,
- flashes of imagination of the sci-fi type,
- strong drive for competence,
- deep moral sense,

which tells us quite a lot.

It tells us that Tom is very bright, does not like to read, as he apparently is not that lyrical in his writing, is not very artistic, has a sense of humor, has a passion for sci-fi, is disciplined, and has a sense of justice[81].

As to the second argument, which is also the hardest to answer. It is taught at the psychology courses, that personality traits tend to be stable over long periods of time. So, if Tom did not like to write in high school[82], than chances are that he would pick a field op specialization in which he would not have to read and write a lot. Which would make humanities and education less probable a field of choice, and computer sciences more probable.

But, lest we forget, the original thesis under discussion was that normative rules of prediction tend to be neglected, as people tend to neglect the base rate; not the alternative thesis that the prediction group should have disregarded the description of Tom, because of the clinical invalidity of projective personality tests and the possibility that Tom might have changed his personality over the course of the few years between high school and college.

If Kahneman and Tversky wish to prove their initial thesis, then at this point, having presented their experimental data, they should proceed to demonstrate that their subjects could not possibly have used the normative, that is, Bayesian, rules of prediction, as those rules would have implied results other than those that were observed. Which they do not.

So, not only do Kahneman and Tversky manage to confuse the participants of the prediction group into thinking that information was given, where in actuality there was none[83]. They also manage to derail their own critical discussion of their actual thesis, by not staying at the issue at hand.

Having come to the end of our discussion of the representative heuristic, we find that the reported plausibility judgments by the prediction group are not inconsistent with a possible use of the Bayesian 'heuristic; (F.26) through (F.29), seeing that the odds (F.27) may be assumed to lie in the realm of the possible. This makes the Kahneman and Tversky experiment inconclusive.

Appendix G. The Framing Effect

The assumption that preferences are not affected by variations of irrelevant features of options or outcomes is called invariance, [59].

According to Kahneman and Tversky invariance is an essential aspect of rationality, which is violated in demonstrations of framing effects, [34]. Now, in order to

[81] Moreover, what is left out also gives us some tentative information on Tom's psychologist.
[82] A liking, admittedly, is not a personality trait, but still.
[83] Or so they believe.

discuss these framing effects, we will first have to discuss the topic of loss and gain adaptation[84].

Imagine a person who is involved in a business venture, who has lost 2000, and now is facing a choice between a sure gain of 1000 and an even chance to win 2000 or nothing, [33]. If he has not adapted to his loss, he is likely to add this loss to all the outcomes and, consequently, by way of the Weber-Fechner law (4.3), code the problem as a choice between the following utility distributions

$$p(u|\, D_1) = \begin{cases} 1.0, & u \;= q \log \frac{S_0 - 1000}{S_0} \end{cases} \tag{G.1}$$

and

$$p(u|\, D_2) = \begin{cases} 0.5, & u \;= q \log \frac{S_0 - 2000}{S_0} \\ 0.5, & u \;= q \log \frac{S_0 - 0}{S_0} \end{cases} \tag{G.2}$$

where S_0 is the pre-loss asset position and q is the Weber constant for money.

Looking at the increments in assets, it is predicted that our business man will tend to prefer D_2 over D_1, as this is the preferred choice under risk seeking in the negative domain[85]. It follows that a failure to adapt to losses may induce risk seeking in the negative domain, [33]. Stated differently, a person who has not made peace with his losses is likely to accept gambles that would be unacceptable to him otherwise. We may find support for this hypothesis by the observation that the tendency to bet on long shots will increase in the course of a betting day, [45].

However, if our business man has adapted to his loss, then he will update his pre-loss asset position S_0 to an adjusted asset position $S_0^{(\text{adj.})}$ in which the loss is discounted:

$$S_0^{(\text{adj.})} = S_0 - 2000$$

and code the problem as a choice between the utility distributions

$$p(u|\, D_1) = \begin{cases} 1.0, & u \;= q \log \frac{S_0^{(\text{adj.})} + \;1000}{S_0^{(\text{adj.})}} \end{cases} \tag{G.3}$$

and

$$p(u|\, D_2) = \begin{cases} 0.5, & u \;= q \log \frac{S_0^{(\text{adj.})} + \;0}{S_0^{(\text{adj.})}} \\ 0.5, & u \;= q \log \frac{S_0^{(\text{adj.})} + \;2000}{S_0^{(\text{adj.})}} \end{cases} \tag{G.4}$$

where $S_0^{(\text{adj.})}$ is the post-loss asset position and q is the Weber constant for money.

Again looking at the increments in assets, we see that the signs of these increments have reversed. By this reversal in the sign of the asset increments, we go from a risk seeking in the negative domain scenario to a risk aversion in the positive domain

[84] What we will call loss and gain adaptation is called *shifts of reference* by Kahneman and Tversky, [33].

[85] See the section on the fifth supporting contact.

scenario[86], [33]. So, it is now predicted that our business man, having already adapted to his loss, will tend to reverse his preferences, and choose D_1 over D_2.

Having introduced the concepts of loss and gain adaptation and the adjusted initial wealth $S_0^{(\text{adj.})}$, we now may turn to the discussion of the framing effect.

Consider the following problems, which were presented to two different groups of subjects, [33].

Group 1: In addition to whatever you own, you have been given 1000. You are now asked to choose between

$$p(O|\,D_1) = \begin{cases} 0.5, & O = 0 \\ 0.5, & O = 1000 \end{cases}$$

and

$$p(O|\,D_2) = \begin{cases} 1.0, & O = 500 \end{cases}$$

Group 2: In addition to whatever you own, you have been given 2000. You are now asked to choose between

$$p(O|\,D_1) = \begin{cases} 0.5, & O = -1000 \\ 0.5, & O = 0 \end{cases}$$

and

$$p(O|\,D_2) = \begin{cases} 1.0, & O = -500 \end{cases}$$

It is found that 84% of $N = 70$ subjects in Group 1 prefer bet D_2 over bet D_1; whereas 69% of $N = 68$ subjects in Group 2 prefer bet D_1 over bet D_2, indicating risk seeking in the negative domain, [33].

This result may indicate that the subjects in both groups adapted to the respective gifts of 1000 and 2000, by adjusting their initial S_0 into a new $S_0^{(\text{adj.})}$, before choosing between the two bets. Since we would have expected to see the same preferences for both groups had the gifts been discounted in the outcomes[87]. But instead, we observe a reversal in preferences. So, we conclude that the gifts must have been discounted in initial wealth, rather than in the outcomes.

Furthermore, based on the unadjusted outcomes alone, as given in the corresponding outcome probability distributions, we would expect both the observed risk aversion in the positive domain in Group 1, that is, the observed preference of D_2

[86] *idem.*

[87] It may be checked that a discounting of the respective gifts in the corresponding outcomes would have resulted in identical outcome probability distributions for both Groups 1 and 2:

$$p(O|\,D_1) = \begin{cases} 0.5, & O = 1000 \\ 0.5, & O = 2000 \end{cases}$$

and

$$p(O|\,D_2) = \begin{cases} 1.0, & O = 1500 \end{cases} .$$

over D_1, and the observed risk seeking in the negative domain in Group 2, that is, the observed preference of D_1 over D_2. This then also points to an adjustment of the initial wealth, rather an adjustment of the outcomes.

However, an alternative, more parsimonious, explanation for this framing effect, or, equivalently, the observed reversal in preferences in both groups, would be that the respective gifts of 1000 and 2000 were neglected by the subjects, [33].

And we would tend to agree with Kahneman and Tversky on this one. For, when reviewing these hypothetical choices, we ourselves overlooked these gifts too.

But where Kahneman and Tversky see the framing effect as an indication that such monetary gifts, as a rule, will not factor into our real-life decisions, we quote, [33]:

> The apparent neglect of a bonus [our gift] that was common to both options [our decisions D_1 and D_2] in Problems 11 and 12 [our Groups 1 and 2] implies that the carriers of value or utility are changes of wealth, rather than final asset positions that include current wealth. This conclusion is the cornerstone of an alternative theory of risky choice [their prospect theory].

We, instead, propose that this neglect of the gifts point to the limitations of the experimental method of hypothetical choices, as employed by Kahneman and Tversky.

Speaking strictly for ourselves, the receiving of a real-life gift of either 1000 or 2000 euros would be quite the momentous occasion. Consequently, we can hardly imagine neglecting such a substantial sum of money, or, for that matter, not factoring its occurrence in our every monetary decision[88].

So, it would seem, at least based on the above [33] quotation, that Kahneman and Tversky build their prospect theory around a phenomenon which, as our introspection would suggest, is nothing but an experimental artifact of the method of hypothetical choices.

[88]The question if such a gift would be discounted into our initial wealth S_0, or into the specific outcomes ΔS of some set of decisions D_i, would be dependent on the particular context in which the gift was received.