On the rationality of decisions with unreliable probabilities

Fernando Birman
Delft University of Technology

Abstract
The standard Bayesian recipe for selecting the rational choice is presented. A familiar example in which the recipe fails to produce any definite result is introduced. It is argued that a generalization of Gärdenfors' and Sahlin's theory of unreliable probabilities — which itself does not guarantee a solution to the problem — offers the best available approach. But a number of challenges to this approach are also presented and discussed.

Keywords
Rationality; Bayesianism; utility; probability; reliability.

Let us suppose that a person called Mary, who likes to carefully and methodically ponder all her decisions, goes to the furniture store to purchase a new couch. And let us suppose, to keep things from getting unnecessarily knotty, that she has trimmed her options down to three. Couch A is very cute and stylish but (not surprisingly) rather uncomfortable and pricey. Couch B is not so stylish, but it is more comfortable than couch A, and its price is average. Couch C is by far the most comfortable of all three couches, and it is a bargain. Its design, however, leaves much to be desired. Mary knows that, depending on whether she settles on design, price or comfort as her priority, her ordinal preference ordering will be different (Table 1). But the problem, of course, is that none of these attributes stand alone before her as the sole priority.

Now, the question of interest in this article is this: What should rational people do in order to solve such a choice problem *rationally*, as opposed to simply rely on chance or blind impulse? What should rational people do, that is, to rationally *reconcile* conflicting preference orderings? I will first introduce the standard Bayesian recipe for selecting the rational choice, which provides nowadays the canonical

Disputatio, Vol. III, No. 26, May 2009
Received on 27 July 2009

answer to this question, and then examine some related but much less studied issues that arise when some of the available choice options involve — more or less — unreliable probabilities.

Design as priority	Couch A ≻ couch B ≻ couch C
Price as priority	Couch C ≻ couch B ≻ couch A
Comfort as priority	Couch C ≻ couch B ≻ couch A

Table 1

The standard Bayesian recipe for selecting the rational choice goes essentially like this. We must first identify those attributes that either increase or decrease the desirability of each single competing option. In our case, I think it is safe to assume that our previous attributes almost invariably determine our choices — other attributes like (say) size, color, and manufacturer's name being equal — when buying a new couch: comfort *M*, design *D*, and price *P*[1]. Second, standard Bayesianism asks us to ascribe cardinal utilities (invariant in their representational properties up to a positive affine transformation) to all individual attributes associated with each single competing alternative. In the simplest of cases, we can assume that such allocations correspond to single real scalars, rather than continuous functions — we will get to the latter case, which is by and large the standard in the context of science and engineering, below[2]. We can assume, for example, that Mary gives couch A, B, and C the following scorecards:

[1] Exactly the same Bayesian procedure, though slightly more intricate, would apply if we let all six or more of these attributes change among competing alternatives. It is also worth noting that none of the attributes, including price, must necessarily be associated with a quantitative value: they can be, for example, assessed in terms of *good*, *fair*, *bad*, and so on, though their utilities (see next) must of course be quantitative.

[2] Only in very simple cases — like the one presented here, or (say) Savage's famous omelette, or Gärdenfors' and Sahlin's Miss Julie and the tennis match — can single attribute utilities be introduced in the form of fixed single scalars. In science and engineering, in contrast, where single attributes are extremely variable (given the

	$u(M)$	$u(D)$	$u(P)$
A	2	5	2
B	4	4	3
C	5	1	4

Table 2

And third, we are required to associate a scaling value, which represents how strongly we want (or dislike) each single attribute against all others, to all individual attributes. Let us assume, in line with the assumption above that none of the attributes are her sole priority, the following scaling values k for M, D, and P according to Mary:

	k
M	4
D	3
P	3

Table 3

Provided this, and provided that we are still dealing with a situation where the subjective probability measure defined over the possible states of the world is assumed to be either 0 or 1 (i.e., a situation

extreme variability of parameters like mass, elasticity, thermal conductivity, and so on, which determine the value of such attributes), single attribute utilities are usually introduced as (often continuous) functions. In order for this article to be of relevance not only to philosophers but to anyone with an interest in decision theory in its most diverse applications, I will consider later on the most general formulation of the issues, of which the aforementioned examples are only special cases.

of epistemic certainty where no other probabilities are involved), we are finally asked by the Bayesian to determine the overall utility of A, B, and C simply by adding all single attribute utility measures, factorized by the corresponding scaling values, allotted by Mary to each individual attribute *M*, *D*, and *P* of A, B, and C.

$U\,(\mathrm{A}) = u(M)_{\mathrm{A}}\, k_{\mathrm{M}} + u(D)_{\mathrm{A}}\, k_{\mathrm{D}} + u(P)_{\mathrm{A}}\, k_{\mathrm{P}} = 8 + 15 + 6 = 29$
$U\,(\mathrm{B}) = u(M)_{\mathrm{B}}\, k_{\mathrm{M}} + u(D)_{\mathrm{B}}\, k_{\mathrm{D}} + u(P)_{\mathrm{B}}\, k_{\mathrm{P}} = 16 + 12 + 9 = 37$
$U\,(\mathrm{C}) = u(M)_{\mathrm{C}}\, k_{\mathrm{M}} + u(D)_{\mathrm{C}}\, k_{\mathrm{D}} + u(P)_{\mathrm{C}}\, k_{\mathrm{P}} = 20 + 3 + 12 = 35$

Table 4

This is, in a nutshell, what the standard Bayesian recipe would ask of rational epistemic agents that, in a context of certainty, are struggling to come up with a resolution to their choice problem[3]. In our present case, this recipe would tell us that, in spite of the fact that A is the maximally stylish couch and C is the maximally comfortable and least expensive couch, the rational option is actually B — the option which maximizes overall utility. Now, it is easy to see, even in such a simple case like this, that much depends on how the single attribute utility measures and scaling values are respectively distributed by the agents. There exist competing lottery methods whose object is precisely to offer an accurate framework for assessing what utility measure and scaling value corresponds to each single attribute, and for aggregating such measures when many epistemic agents are involved[4].

[3] Standard Bayesianism stands here for either a causal or evidential decision theory as famously presented, for example, in von Neumann & Morgenstern 1944, Luce & Raiffa 1957, Savage 1972, Jeffrey 1983, and Skyrms 1990.

[4] The possibility of aggregating individual utilities is often countered with skeptical arguments — such as the 'zero-point utility' and 'unit of utility' arguments — which maintain, roughly, that inter-subjective comparisons of utilities are, in general, meaningless — they claim, in a word, that no co-ordinatization of the utility-space, and no metric system for this space, can be *objectively* singled out among all possible options. It is such reluctance to consider inter-subjective comparisons of utilities, explicitly rejected by Arrow, what makes the Impossibility

Before we illustrate this point, let me introduce the most general (under epistemic certainty) form of this Bayesian recipe, whereby any finite number *n* of attributes is involved, and all single attribute utilities are continuous functions rather than scalars.

For independent single utilities, the overall utility of each competing choice (in a context, again, of certainty) is determined from[5]:

$$[1] \quad U(X) = \frac{1}{K}\left\{\left[\prod_{i=1}^{n}\left(Kk_i u_i(x_i)+1\right)\right]-1\right\}$$

where X is the vector of single attributes x_1, x_2, …, x_n associated with the choice option in question, u_i (x_i) is the single attribute utility function for attribute x_i, k_i is the scaling value of single attribute x_i, and K is a normalizing constant derived from k_i, where $1 + K = \prod (1 + Kk_i)$.

There exist, as noted, distinct lottery methods for plotting the values of u_i (x_i) and k_i. One such method consists, roughly, in asking all agents to imagine that two competing options are being considered, each alike in every respect except one: the attribute level x_j for the 'certain' option is known with certainty to be (say) 24 while x_j for the 'lottery' option has (say) a 60% probability p of being $x_{j\ max} = 30$ (maximal utility) and a 40% probability $1 - p$ of being $x_{j\ min} = 20$ (minimal utility)[6]. The agent is then asked: 'Which option do you prefer: the certain option, the lottery option, or are you indifferent?'. If she responds 'The certain option' the value of x_j is decreased (increased) to a less desirable level that is half-way between 24 and 20 (or half-way between 24 and 30, if x_j's desirability goes down when x_j goes up). The agent is again required to express her preference, and if the lottery option is now preferred, the value of x_j is increased until the decision maker is indifferent between the 'certain' option and the 'lottery' option. A point on the single attribute utility function u_j (x_j) is plotted as $p\ u_j\ (x_{j\ max}) + (1 - p)\ u_j\ (x_{j\ min})$ when such indifference

Theorem's disturbing results in social choice theory a demanding challenge for the Bayesian — see Arrow (1950; 1963), Sen 1970, Suzumura 1983.

[5] I follow here, and later, Keeney & Raiffa 1976, chapter 6.

[6] As noted before, attributes must not necessarily be quantitative. The same procedure (though slightly more cumbersome, given that scalars often provide a more extensive and fine-grained spectrum of choices) would apply if the relevant attributes were evaluated in terms of a qualitative scale, like (say) the descriptor scale mentioned before: *excellent*, *good*, *fair*, and so forth.

points are reached. And the procedure is iterated for different *p*s until the entire function, or at least the function within some interval of interest, is sufficiently outlined. Similar types of lottery questions are employed for plotting the levels of k_i[7].

One more technical ingredient before we get to our problem.

So far we have been dealing with a choice problem in a context of certainty. It is common in the decision-theoretic literature, despite frequent terminological discrepancies that still muddle many discussions, to distinguish between decision problems in a context of: (i) certainty, (ii) risk, (iii) uncertainty and (iv) ignorance. A choice problem under risk is assumed to involve non-trivial (i.e., other than 0 and 1) subjective probability measures over the possible states of the world. And such probability measures are taken to be fully reliable. Choice problems under uncertainty and ignorance, in contrast, are considered to be, respectively, those whereby the probability distributions are either unreliable — and the measure of such unreliability is determined differently in different theories of unreliable probabilities[8] — or completely unknown. I will now introduce a slightly different decision problem regarding couch A, B and C. And I will incorporate probabilities into it. In doing so, I will set forth the problem involving unreliable probabilities that I want to discuss in the remainder of this article. But let us first upgrade equation [1]. For independent single utilities, the overall *expected* utility of each alternative choice is determined from:

$$[2] \quad E[U(X)] = \frac{1}{K}\left\{\left[\prod_{i=1}^{n}\left(Kk_i\left(\int_{x_i \min}^{x_i \max} u_i(x_i) f_i(x_i) dx_i\right)+1\right)\right]-1\right\}$$

where everything is as before, and f_i (x_i) is the probability density function correlated with attribute x_i. For all choice problems under

[7] See Keeney & Raiffa 1976, chapters 4 and 5. For a detailed description of one of the first computational models developed to carry out these procedures, see Thurston 1991.

[8] Three of the major theories are due to Ellsberg 1961, Levi (1974; 1980; 1982; 1986), and Gärdenfors & Sahlin 1982*a*, reprinted in 1988. Earlier theories can be found in Wald 1950 and Hurwicz 1951. And alternative approaches can be found in Skyrms 1980, Kaplan 1983, Sahlin 1983, and Baron 1987. For more recent theories, with emphasis on the implementation in science and technology and their main focus on *fuzzy sets*, see Dubois & Prade 1988, Dubois, Prade & Yager 1996, Antonsson 2001.

risk, $f_i(x_i)$ is assumed to be completely reliable. But for problems under uncertainty (and, of course, ignorance), this assumption does not hold. How are these probabilities obtained in the first place? Bayesian decision theory is based on subjective probabilities. So we start out with a set probability measures defined (in principle) over all possible states of the world. Such initial probability measures may indeed be arbitrary. But new information — arising, for example, from new observations — will adjust the original measures so that they converge toward an 'objective' distribution, as established by de Finetti's famous Representation Theorem[9]. In practical applications of Bayesianism, it is also routine to supplement such so-called 'judgmental information' with statistical estimates[10], for which different methods exist which determine *goodness-of-fit* of probability distributions given the available statistical data — the most recognized of such methods being *Chi-square*, *Kolmogorov-Smirnov* and *Anderson-Darling*. A problem with the statistical approach, however, which compromises the probabilistic assignments the Bayesian makes from hard data, is that different *goodness-of-fit* methods occasionally prescribe different probability distributions to exactly the same data. This is referred to in the literature as the 'distribution *arbitrariness* problem'[11]. In this paper, I will just ignore this problem.

Let us get back now to the furniture store example.

Imagine that Mary, following the standard Bayesian recipe, has finally settled on couch B. Now she faces the final, and maybe most

[9] See de Finetti 1937, reprinted in Kyburg & Smokler 1964. The concept of 'objective' probability can be literally understood, of course, as a reflection of actual propensities in the world or — and this is the view that de Finetti would favour — an inter-subjective probability distribution to be obtained while asymptotically reaching the end of inquiry.

[10] See for example Ang & Tang 2007, where it is argued that Bayesianism is the best tool for dealing with probabilistic phenomena, since it provides, unlike the classical probabilistic methods based exclusively on statistical data, a framework for easily combining scientific and technological *expert knowledge* with brute statistical information.

[11] See Ditlevsen 1994. See also Hansson 2009. It is worth noting that a better name would be, I believe, 'distribution *under-determination* problem', for the problem seems to be rather the under-determination of probability distributions by data. And under-determination, I think, does not necessarily entail arbitrariness: we may have good *non-arbitrary* reasons for accepting Q rather than ¬Q, although no reason can be given which *determines* Q.

challenging, decision problem. She is informed that couch B is manufactured at two different locations: Ψ-town and Φ-town. It is of the utmost importance that couch B be delivered precisely on time — say, June 1st — for Mary is about to embark on a very long trip — on, say, June 2nd — and the couch must be necessarily in place *before* she departs. She is informed, in addition, that the factory in Ψ-town delivers on time — statistically — 50% of the time. That is, provided a long sample of cases (seventy years of work, with hundreds of deliveries per year), roughly 50% of those deliveries were right on time[12]. Finally, Mary is informed that the factory in Φ-town delivers on time 100% of the time when they do not have old pending orders but 0% of the time — that is, no timely deliveries at all — whenever there is a backlog[13]. And there is no available information whatsoever regarding whether they presently have a backlog or not — and, as most readers familiar with furniture dealers and manufacturers will understand, they are not at all responsive to phone calls or emails. Mary has, yet again, settled on couch B. So she must now choose *rationally* — and by that we mean that she must choose so as to promote her goals, for we still follow here the old *Humean* notion of instrumental rationality inherent to Bayesianism[14] — in such a way that, given all she knows, couch B has the most chances of being delivered on time.

[12] It would be of course unrealistic if I said that *exactly* 50% of deliveries were right on time. Readers who will find this unsettling for the argument that comes next may well assume that that was indeed the case.

[13] Due to, say, returned merchandise in need of repair. Otherwise, backlogs would be impossible in Φ-town. On the other hand, when the factory is behind schedule, new employees are hired to clear the backlog.

[14] Hume famously stated that the fixation of goals falls outside the province of rationality, and thus it is not contrary to reason '(…) to prefer the destruction of the whole world to the scratching of my finger' (1969: 463). The object of rationality, rather, is to tell any rational individual how to achieve whatever goal she has set her mind on. As Simon put it, '(…) reason is wholly instrumental. It cannot tell us where to go; at best it can tell us how to get there. It is a gun for hire that can be employed in the service of whatever goals we have, good or bad' (1983: 7–8).

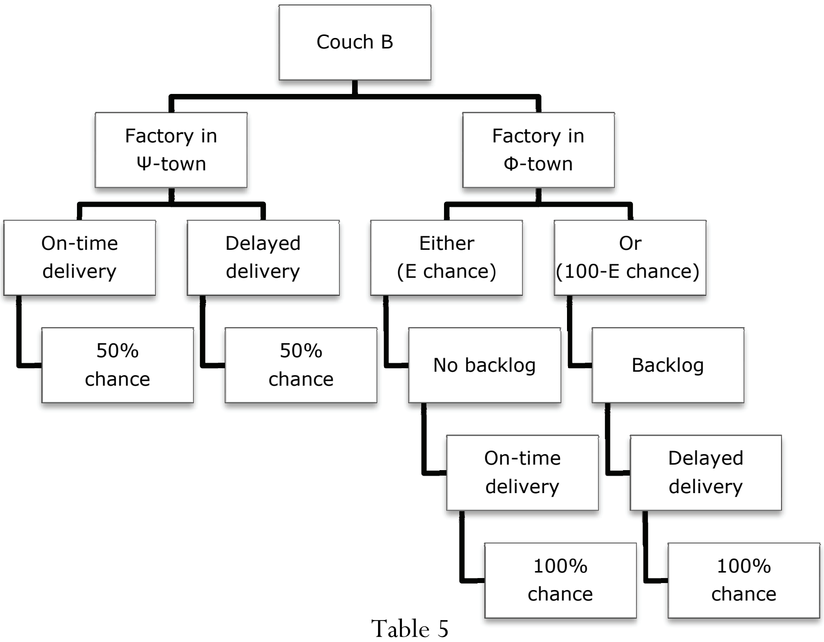

Table 5

What should Mary do? What is the *rational* solution to the problem? According to the standard Bayesian, these two situations — placing the order in Ψ-town or Φ-town — are, from a decision-theoretic perspective, indiscernible (Table 5). The probability of getting the couch delivered on time when the order is placed in Ψ-town is, given the available statistical information, 0.5. And the subjective probability defined over the two possible states of the world that correspond to the situation of placing the order in Φ-town — namely, there is a backlog or there is no backlog — is also, for each state, 0.5. That is, provided that Mary possesses no information whatsoever regarding whether the factory in Φ-town has presently a backlog or not, and provided that contextual information concerning couch factories suggests that there is no reason to believe that one state of affairs is in principle more probable than the other (there is no reason to believe that the value of E in Table 5 is anything other than 50), the most sensible way to allocate initial subjective probabilities — which would, of course, be susceptible to change if more information became available — goes like this: (i) 0.5 probability that there is a backlog and, accordingly, that couch B is not delivered on time; (ii) 0.5 probability that there is no backlog and, hence, that couch B is delivered on time. These two situations, then, are epistemically indiscernible for the Bayesian. But it is clear that the situations are not evenly appealing in terms of their subjective *potential* at promoting the desired goal — getting the couch delivered on time. They are not equally appealing, in other words, in terms of their instrumental rationality. The standard Bayesian, although she helped us before, cannot help us now.

This problem, which goes back at least to Peirce[15] and to Popper's and Ellsberg's famous paradoxes[16], does not yet have a universally accepted solution. I will now briefly offer my own solution, which is a generalization of Gärdenfors' and Sahlin's well-known theory of unreliable probabilities — which itself, given its reliance on *minimal* satisfactory levels of epistemic reliability, which in turn determine the set of probability distributions from which the MMEU criterion maximizes minimal expected utility, does not guarantee a solution to the problem[17]. And I will then say why this solution, which (to my mind) is the best available approach, may not yet be entirely satisfactory.

[15] Peirce (1932: 421) famously said: '(…) to express the proper state of belief, not *one* number but *two* are requisite, the first depending on the inferred probability, the second on the amount on knowledge on which that probability is based'. This second number, in this case, would presumably break the tie for a Bayesian.

[16] See Popper (1959: 407–408) for a standard formulation of the so-called 'paradox of ideal evidence', although the paradox was discovered by Peirce himself. It shows, in a word, that a strict Bayesian should be as inclined to believe that the next toss of a fair coin is going to yield heads as she is to believe that the next toss of an entirely unknown coin is going to yield heads, which is evidently unwarranted. And see Ellsberg (1961: 653–654) for the introduction of his famous paradox, to which we will return later.

[17] If, for example, the degrees of reliability of the 50% chance of timely delivery from Ψ-town and Φ-town both fell below such minimal satisfactory level of reliability, then Gärdenfors' and Sahlin's theory would as well be silent with respect to what choice, either placing the order in Ψ-town or Φ-town, is to be preferred. This minimal satisfactory level of reliability, still, could be viewed as a *fluid* parameter, so that the *context* dictates its value — and it conveniently does so in a way that it is always below the threshold of reliability of some probability. But the catch now would be, I think, that the resulting theory is hardly normative, for we now allow this parameter to be a mere reflection of an agent's varying risk appetites (more on this in footnote 23). The reader must remember that Gärdenfors' and Sahlin's decision theory consists of two central moments. First, the decision maker must restrict, provided all competing alternatives in a decision situation, the set of possible probability measures for that situation 'to a set of probability measures

I believe, in a word, that the rational response to this choice problem — whether to place the order in Ψ-town or Φ-town — is to place the order in Ψ-town. Why? Because the 50% chance of getting the couch delivered on time that follows from this choice is more *reliable* (more information, of statistical or judgmental nature, is behind this probabilistic appraisal) than the 50% chance obtained from placing the order in Φ-town. Let us define δ as the real-valued degree of reliability, with $0 \leq \delta \leq 1$, that corresponds to all subjective probability distributions defined over the possible states of the world. In our present case, this amounts to the contention that the 50% chance of timely delivery associated with the choice of placing the order in Ψ-town has a δ_1 measure of reliability, and the 50% chance associated with placing the order in Φ-town has a not necessarily identical δ_2 measure of reliability. And so my previous claim concerning the 50% chance of timely delivery from Ψ-town being more reliable than exactly the same chance of timely delivery from Φ-town amounts now to the simple claim that $\delta_1 > \delta_2$.

What must be shown at this point is that, all other things being equal, higher δ's correspond to higher expected utilities. But is this *necessarily* the case? My discomfort with the present approach, which has just been insinuated, comes from the realization that, although this is *usually* the case, it might not *always* be the case. But let us first explore how higher δ's may correspond at all to higher expected utilities.

The original overall utility of choosing couch B, as opposed to couches A or C, is:

$$[3] \quad U(X_B) = \sum_{i=1}^{n} k_i u_i(x_i)$$

where X_B is the vector of single attributes $x_1, x_2, \ldots, x_n$ of choice B. No consideration as to the on-time delivery of couch B is yet included among these attributes.

The overall expected utility of opting for couch B when couch B is manufactured in Ψ-town is:

with a 'satisfactory' degree of epistemic reliability' (1982*a*: 369). Second, the decision maker must apply within this restricted set of probability measures the *maximin* criterion for expected utilities: she must choose that alternative which has the largest minimal expected utility. For a short description of the theory, see Gärdenfors & Sahlin 1982*a*, reprinted in 1988. Criticism can be found in Levi 1982, with a reply from the authors (Gärdenfors & Sahlin 1982*b*), and Levi 1985.

[4] $$E[U(X_B + x_\Psi)] = \sum_{i=1}^{n} k_i u_i(x_i)\Delta_i$$

where probability $\Delta_i = 0.5$ if $x_i = x_\Psi$ and $\Delta_i = 1$ otherwise. The additional attribute of on-time delivery is here represented by x_Ψ. The single attribute utility u_Ψ (x_Ψ) and scaling value k_Ψ are both normalized at 1.

The overall expected utility of choosing couch B if manufactured in Φ-town is:

[5] $$E[U(X_B + x_\Phi)] = \sum_{i=1}^{n} k_i u_i(x_i)\Delta_i$$

where probability $\Delta_i = 0.5$ if $x_i = x_\Phi$ and $\Delta_i = 1$ otherwise. The additional attribute of on-time delivery is here represented by x_Φ. The single attribute utility u_Φ (x_Φ) and scaling value k_Φ are both normalized, again, at 1.

Now, the reader can easily verify that equations [4] and [5] will yield exactly the same result. The utility and scaling values assigned to the attribute of on-time delivery are the same in [4] and [5]. And the probability allocated to such event is also the same (i.e., 0.5) in both equations. And all the remaining parameters are, naturally, the same[18]. This is the reason why these two situations are precisely the same for the standard Bayesian. And this is the reason why it was contended above that the standard Bayesian cannot possibly break the tie between them: for her, these two situations are epistemically indiscernible.

My proposal to break the tie, and therefore to be able to rationally decide between placing the order in Ψ-town or Φ-town, is to incorporate Gärdenfors' and Sahlin's value δ into the equations, although not as a measure of minimal satisfactory degrees of epistemic reliability, which are in turn responsible for artificially setting a cap on what probabilities will actually make it to the utility calculus, but rather as a measure of the reliability of all subjective probabilities defined over the possible states of the world, for epistemic agents a_1, a_2,…, a_m at

[18] More precisely, all k_i except for k_Ψ in [4] and k_Φ in [5] are 0 at this stage, if we expect the attribute of on-time delivery to be *necessary*, as stipulated above. I thank one of the referees for pointing this out to me.

time t[19]. When choosing couch B if manufactured in Ψ-town, we obtain as measure of the overall expected utility:

$$[6]\quad E\left[U\left(X_B + x_\Psi\right)\right] = \sum_{i=1}^{n} k_i u_i\left(x_i\right)\Delta_i \delta_i$$

where everything is as before and $\delta_i = \delta_1$ if $x_i = x_\Psi$ and $\delta_i = 1$ otherwise.

The overall expected utility of choosing couch B, on the other hand, when couch B is manufactured in Φ-town is:

$$[7]\quad E\left[U\left(X_B + x_\Phi\right)\right] = \sum_{i=1}^{n} k_i u_i\left(x_i\right)\Delta_i \delta_i$$

where everything is as before and $\delta_i = \delta_2$ if $x_i = x_\Phi$ and $\delta_i = 1$ otherwise. But we assumed earlier that $\delta_1 > \delta_2$, for there is statistical information behind Δ_Ψ and no information at all, other than the assumption that backlogs are neither more likely nor less likely in Φ-town than no backlogs, behind Δ_Φ. It follows immediately from this, as readers can confirm by themselves, that $E[U(X_B + x_\Psi)] > E[U(X_B + x_\Phi)]$. The rational option is, therefore, to place the order in Ψ-town.

The most general result here, for any finite number n of attributes $x_1, x_2, \ldots, x_n$ of a choice T open to agents $a_1, a_2, \ldots, a_m$ and displaying single independent utility functions $u_i(x_i)$, with probability density functions $f_i(x_i)$ and reliabilities δ_i between 0 and 1, is:

$$[8]\quad E\left[U\left(X_T\right)\right] = \frac{1}{K}\left\{\left[\prod_{i=1}^{n}\left(Kk_i\left(\int_{x_i \min}^{x_i \max} u_i\left(x_i\right) f_i\left(x_i\right)\delta_i dx_i\right)+1\right)\right]-1\right\}$$

The old rule of maximizing expected utility will then determine, as it does in the standard case, that T must be chosen as the *rational* option

[19] This proposal aims not only at capturing the way scientists and engineers deal in reality with probabilistic knowledge, but also providing a decision theory that can be effectively used in everyday practice. It is also worth noting, incidentally, that the analysis presented in this work is strictly static. A fully *dynamic* analysis ought to consider how the reliability measures are conditionalized on the amount and quality of information as it becomes available to the epistemic agents. The δ's presented here, therefore, are just fixed synchronic slices of an evolving parameter.

if and only if no other available option has either equal or higher overall expected utility[20].

I will now turn to address two concerns that the approach presented in this article immediately brings to mind. The first, which I think is less of a challenge, has to do with the question of how to determine the reliability measure of our probabilities. How is it, in other words, that we can declare with confidence, given what we know, that (say) $\delta_1 > \delta_2$? And the second concern, insinuated before, has to do with the question of whether higher reliabilities —appraised by whatever means we turn to rely on — do necessarily correspond (all other things being equal) to higher expected utilities. As for the first question, I believe this version of Bayesianism must be explicitly based on a subjective notion of reliabilities (and probabilities), exactly like standard Bayesianism is based on subjective probabilities. It is, I think, for scientists and engineers, and whoever employs the theory, to assess how reliable their probabilities are. And there is in principle no constraint on how to specify such subjective measures[21]. A difficulty though, which also threatens the second-order-probabilities approach to the 'reliability' of probabilities, is that this opens the door to an infinite regress problem: to assess how reliable our measures of reliability are, we need a further measure of reliability gauging how reliable our measures of reliability are, and so on. All epistemic inquiries, however, must stop somewhere, even those of a scientific and technological nature — and more so those involving the trivial task of buying furniture. And, I think, uncovering how reliable our probabilities are looks like a reasonable stopping point. However, decision makers under special circumstances may very well be pragmatically inclined to push further on, and this theory does not preclude them from doing so.

[20] There are, of course, different decision criteria for dealing with uncertainty and ignorance — like *maximin*, *maximax*, *minimax regret*, Ellsberg's rule, Levi's test — that transcend the maximizing expected utility rule. In the present theory, however, the only fundamental decision criterion that we need, which may indeed be supplemented by other decision criteria under special circumstances, is that of maximizing overall expected utility.

[21] It is likely, for example, that this subjective assessment of reliabilities will be as involved with heuristics and biases as the subjective assessment of probabilities. For a famous analysis of the latter case, see Kahneman, Slovic & Tversky 1982.

The second problem, I think, is harder to address. It seemed at first obvious to me that the more you know about your probabilities, the better. However, after presenting the ideas contained in this paper to different audiences, it dawned on me that lower δ's might very well correspond to what people often judge as higher expected utilities. When asked about whether you prefer to toss a fair coin (50% chance of winning) or bet on one of two completely unknown horses running against each other (50% chance of winning, for you allocate your subjective probabilities symmetrically between these two entirely unknown horses), people usually respond: 'I'd rather bet on the horses', though betting on the fair coin has evidently the highest δ[22]. This result, needless to say, falls short of constituting a scientific poll. And I feel reluctant, of course, to draw any conclusions based on anecdotal evidence. However, I find this strange. One may be tempted to argue here that *rationality* is not a descriptive but normative concept, and so those who prefer a bet on the horses to a bet on the fair coin are acting irrationally[23]. While I agree with this view of rationality[24], I find this response unconvincing. Not only those who revealed an inclination for betting on the horses were (I

[22] This question obviously stipulated — though it would be interesting to see how people respond if different money prizes are involved — that both betting on the fair coin and betting on one of the horses pays exactly the same: you get your money doubled if you win and you lose everything if you lose. And, indeed, it was also stipulated that no track record of the horses was available either.

[23] The theory presented in this paper is normative in nature, while it struggles to capture what we intuitively associate with a rational choice. At the same time, it presupposes that able human beings, as a norm, choose rationally *most of the time*. It is worth pointing out that Gärdenfors' and Sahlin's theory, on the other hand, is less of a normative decision theory. The measure of epistemic reliability δ is for them a reflection of how risk averse — or risk seeking — a decision maker is, which determines what probability distributions out of all possible distributions will be part of the MMEU process. And this, of course, determines what decision gets singled out as the rational choice. In this theory, on the contrary, the measure δ, though related to a decision maker's risk attitude, for she may be more or less willing in principle to see epistemic merit in a probability distribution, is viewed rather as a subjective measure of how *robust* probabilities are, given all the available data. So the emphasis is not on how agents handle probabilities, but on what probabilities dictate to agents.

[24] And, despite the skeptical arguments presented by Broome 2007, I see the normativity of rationality as a corollary of the instrumental nature of the latter.

venture to say) exemplars of the *quintessence* of human rationality, but also the contention itself that preferring in this case a bet on the horses is necessarily irrational is, I think, quite feeble.

Another plausible response, of course, to this puzzling situation may go along the following lines. The reliability measure δ of our subjective probabilities over the possible states of the world is a relevant factor in the assessment of expected utilities when certain pragmatic interests and concerns (such as, for example, keeping our families safe, making profits while investing our money, allocating very limited resources to scientific projects, designing reliable products and services for the customer) are being pursued but not when other pragmatic interests and concerns (such as having an enjoyable time while gambling, which does not necessarily imply making the most profit) are at stake. While, again, this looks like a plausible answer, I think that there is still something a bit unsettling about it. Are people in this example really choosing, out of two symmetrical options (i.e., the two-sided coin and the two racing horses) the one with less reliable probabilities? What if they were asked now whether, for equal prizes, they prefer to toss a fair coin or an unchecked coin? Now all the emotions we associate with horse racing are entirely eliminated. Would that make a difference? I am inclined to believe that it would — that is, people would rather bet now on the fair coin (50% chance of winning, with higher δ) over the unchecked coin (50% subjective chance of winning, with a lower δ). So, perhaps, it is not that people are willing to rationally gamble, given two or more options displaying the same probabilities, on that option which involves less reliable probabilities, as is the case of one of the horses against heads, or tails, of the fair coin, but rather that these options are not, to begin with, truly symmetrical: their utilities, besides the identical monetary prizes, are not the same.

It could also be argued here that, in any case, such strange behavior (if, of course, it is to be empirically validated) merely applies to situations where all stakes are small. In other words, it could be contended that people may occasionally be willing to bet on (say) unknown horses rather than a fair coin (all other things being equal) simply because there is not much at stake in those situations. Although this remark is, quite likely, true, I think that it does not really touch on our problem at this juncture: Can probabilities with lower δ's be *rationally* preferred (all other things, again, being equal) to the same probabilities with higher δ's? I believe, given the instrumental notion of rationality with which we are dealing here (for which, recall from before, the scratching of my finger cannot be of more or less import than the destruction of the entire world, as far as rationality is concerned), that any positive response, big or small, would certainly be of significance. Thus, maybe, even if the example above concerning horse racing and a fair coin is far from conclusive, the question still remains whether rationality is *necessarily* entangled, because of its own nature, with higher epistemic reliabilities (all other things, once again, being equal).

I will conclude with a brief description of an ongoing empirical study that, in my view, casts this idea into question[25]. Imagine a choice problem identical in all respects to Ellsberg's famous problem. An urn contains 30 red balls and, in unknown proportion, 60 black and yellow balls. One ball is taken at random. And you get to choose between two possible gambles: (i) you receive $ 10.000 if a red ball is taken and nothing otherwise; (ii) you receive $ 10.000 if a black ball is taken and nothing otherwise[26]. We leave here aside the other gambles that Ellsberg discusses in his work. The most popular gamble between (i) and (ii) among those who volunteer for the test is, famously, (i). And this is consistent with the idea that, all other things being equal, more reliable probabilities are by and large preferred, as a matter of fact, over less reliable probabilities.

Now, let us imagine that we let our volunteers know that the name of this game is 'the biggest loser', loosely based on the weight loss TV show. The immediate goal now is not to win but to lose — volunteers are asked to make the least money after ten rounds, and the biggest loser wins. Remember, once again, that rationality is not about the fixation of goals: it is blind as far as they are concerned. A choice cannot be deemed irrational on the simple notion that it aims at losing rather than winning. What counts is whether it gets us there, be that winning or losing. So we ask our volunteers to *lose* at each of

[25] This research is being conducted by Maarten Franssen, Laurens Rook and myself among Delft University of Technology students. While Franssen and Rook are actively involved in the empirical work and ensuing discussion (forthcoming in print), the opinions presented in this paper do not necessarily reflect their views.

[26] It was found that responses followed a much less clear pattern if volunteers were asked about money prizes that involved, as in Ellsberg's original paradox, smaller dollar stakes. It must be remembered, however, that $ 100 was not such a small stake in the late fifties, when Ellsberg conducted his tests.

the ten rounds — as people often lose at games not to embarrass friends, or to make their children happy. The subjective probability measure defined over the state of the world that corresponds to our randomly taking a red ball under gamble (i) is 1/3. And the same probability, though with a lower δ (for we have no reasons to believe that black and yellow balls are squarely distributed), corresponds to the state of the world whereby we randomly take a black ball under gamble (ii). This is why gamble (i) is generally preferred to gamble (ii)[27], and why decision theories usually advise us to prefer the former over the latter. But the subjective probability measure defined over the state of the world that corresponds to our randomly *not* taking a red ball under gamble (i) is 2/3. And the same probability, but with a lower δ, corresponds to the state of the world whereby we randomly *not* pick up a black ball under gamble (ii). So people who play 'the biggest loser' should still prefer, and most decision theories for uncertainty and ignorance — including the one presented in this paper — would advise them to do so, gamble (i) over gamble (ii).

Preliminary results suggest that people who play this game unmistakably favour, in their rational attempt to lose, gamble (ii) over gamble (i). That is, they unmistakably favour a bet on a black ball being randomly selected over a bet on a red ball being randomly selected. This can be either interpreted as suggesting that those persons, despite all appearances, are not being truly rational, or that rationality is not *necessarily* accompanied by a taste for higher δ — that is, a taste for more reliable probabilities. In other words, empirical results appear to suggest that, if we accept that people playing this game are on the whole deciding rationally, then rational decisions may occasionally favour, all other things being equal, less reliable probabilities.

Fernando Birman
Delft University of Technology
Department of Philosophy
Delft, the Netherlands
f.a.birman@tudelft.nl

[27] Assuming, of course, that other possible probability distributions concerning black and yellow balls (say, 2/3 subjective probability that we randomly take a black ball, and 0 subjective probability that we randomly take a yellow ball) have still much lower reliabilities. Otherwise, agents may well rationally prefer, as some do, gamble (ii) over gamble (i).

References

Ang, Alfredo & Wilson Tang. 2007. *Probability Concepts in Engineering*. New York: Wiley.

Antonsson, Erik. 2001. *Imprecision in Engineering Design*. Pasadena: California Institute of Technology.

Arrow, Kenneth. 1950. A Difficulty in the Concept of Social Welfare. *Journal of Political Economy* 58 (4): 328–346.

Arrow, Kenneth. 1963. *Social Choice and Individual Values*. New Haven: Yale University Press.

Baron, Jonathan. 1987. Second-Order Probabilities and Belief Functions. *Theory and Decision* 23: 25–36.

Broome, John. 2007. Is Rationality Normative? *Disputatio* 23 (2): 161–178.

de Finetti, Bruno. 1937. La Prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 7 (1): 1–68.

Ditlevsen, Ove. 1994. Distribution Arbitrariness in Structural Reliability. In *Structural Safety and Reliability*, edited by G. Schueller, M. Shinozuka & J. Yao. Leiden: Balkema.

Dubois, Didier & Henri Prade. 1980. *Fuzzy Sets and Systems*. New York: Academic Press.

Dubois, Didier, Henri Prade & Ronald Yager. 1996. *Fuzzy Information Engineering*. New York: Wiley.

Ellsberg, Daniel. 1961. Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics* 75: 643–669.

Gärdenfors, Peter & Nils-Eric Sahlin. 1982*a*. Unreliable Probabilities, Risk Taking, and Decision Making. *Synthese* 53 (3): 361–386.

Gärdenfors, Peter & Nils-Eric Sahlin. 1982*b*. Reply to Levi. *Synthese* 53 (3): 433–438.

Gärdenfors, Peter & Nils-Eric Sahlin. 1988. *Decision, Probability, and Utility*. Cambridge: Cambridge University Press.

Hansson, Sven Ove. 2009. Risk and Safety in Technology. In *Philosophy of Technology and Engineering Sciences*, edited by W. Meijers, D. Gabbay, P. Thagard & J. Woods. Forthcoming.

Hume, David. 1969. *A Treatise of Human Nature* (1739–40). London: Penguin.

Hurwicz, Leo. 1951. Some Specification Problems and Applications to Econometric Models. *Econometrica* 19: 343–344.

Jeffrey, Richard. 1983. *The Logic of Decision*. Chicago: University of Chicago Press.

Kahneman, Daniel, Paul Slovic & Amos Tversky. 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kaplan, Mark. 1983. Decision Theory as Philosophy. *Philosophy of Science* 50: 549–577.

Keeney, Ralph & Howard Raiffa. 1976. *Decisions with Multiple Objectives*. New York: Wiley.
Kyburg, Henry & Howard Smokler. 1964. *Studies in Subjective Probability*. New York: Wiley.
Levi, Isaac. 1974. On Indeterminate Probabilities. *Journal of Philosophy* 71: 391–418.
Levi, Isaac. 1980. *The Enterprise of Knowledge*. Cambridge (MA): MIT Press.
Levi, Isaac. 1982. Ignorance, Probability and Rational Choice. *Synthese*, 53 (3): 387–417.
Levi, Isaac. 1985. Imprecision and Indeterminacy in Probability Judgment. *Philosophy of Science* 52: 390–409.
Levi, Isaac. 1986. *Hard Choices*. Cambridge: Cambridge University Press.
Luce, Robert & Howard Raiffa. 1957. *Games and Decisions*. New York: Wiley.
Peirce, Charles Sanders. 1932. *Collected Papers*. Edited by C. Hartshorne & P. Weiss. Cambridge (MA): Belknap Press.
Popper, Karl. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.
Sahlin, Nils-Eric. 1983. On Second Order Probabilities and the Notion of Epistemic Risk. In *Foundations of Utility and Risk Theory with Applications*, edited by B. P. Stigum & F. Wenstop. Dordrecht: Reidel.
Savage, Leonard. 1972. *The Foundations of Statistics*. New York: Wiley.
Sen, Amartya. 1970. *Collective Choice and Social Welfare*. San Francisco: Holden-Day.
Simon, Herbert. 1983. *Reason in Human Affairs*. Stanford: Stanford University Press.
Skyrms, Brian. 1980. Higher Order Degrees of Belief. In *Prospects for Pragmatism: Essays in Honor of F. P. Ramsey*, edited by D. Mellor. Cambridge: Cambridge University Press.
Skyrms, Brian. 1990. *The Dynamics of Rational Deliberation*. Cambridge (MA): Harvard University Press.
Suzumura, Kotaro. 1983. *Rational Choice, Collective Decisions, and Social Welfare*. Cambridge: Cambridge University Press.
Thurston, Deborah. 1991. A Formal Method for Subjective Design Evaluation with Multiple Attributes. *Research in Engineering Design* 3 (2): 105–122.
von Neumann, John & Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
Wald, Abraham. 1950. *Statistical Decision Functions*. New York: Wiley.