Decision Under Normative Uncertainty

Franz Dietrich, Brian Jabarian

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2018 – 46

Decision Under Normative Uncertainty

Franz Dietrich
Brian Jabarian

JEL Codes:
Keywords :

PARIS-JOURDAN SCIENCES ECONOMIQUES
48, BD JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 80 52 16 00=
www.pse.ens.fr

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES
ÉCOLE DES PONTS PARISTECH – ECOLE NORMALE SUPÉRIEURE
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE – UNIVERSITE PARIS 1

Decision Under Normative Uncertainty

September 2018[1]

Franz Dietrich
Paris School of Economics & CNRS

Brian Jabarian
U. Paris 1 & Paris School of Economics

Abstract

How should we evaluate options when we are uncertain about the correct standard of evaluation, for instance due to conflicting normative intuitions? Such 'normative' uncertainty differs from ordinary 'empirical' uncertainty about an unknown state, and raises new challenges for decision theory and ethics. The most widely discussed proposal is to form the expected value of options, relative to correctness probabilities of competing valuations. But this meta-theory overrules our beliefs about the correct risk-attitude: it for instance fails to be risk-averse when we are certain that the correct (first-order) valuation is risk-averse. We propose an 'impartial' meta-theory, which respects risk-attitudinal beliefs. We show how one can address empirical and normative uncertainty within a unified formal framework, and rigorously define risk attitudes of theories. Against a common impression, the classical expected-value theory is not risk-neutral, but of hybrid risk attitude: it is neutral to normative risk, not to empirical risk. We show how to define a fully risk-neutral meta-theory, and a meta-theory that is neutral to empirical risk, not to normative risk. We compare the various meta-theories based on their formal properties, and conditionally defend the impartial meta-theory.

1 Normative versus empirical uncertainty

Suppose we must evaluate some choice options. The criterion could take several forms, for instance personal well-being, 'utility' in a suitable sense[2], moral value, social welfare,

[1] We thank various colleagues for stimulating exchanges and in some cases extensive comments. We wish to mention (in alphabetic order) Ron Aboodi, Roland Bénabou, David Black, Richard Bradley, Ryan Doody, David Enoch, Marc Fleurbaey, Hilary Greaves, Alan Hajek, Sergiu Hart, Seth Lazar, François Maniquet, Ittay Nissan-Rozen, Katie Steele, Jean-Marc Tallon, Christian Tarsney, Aron Vallinder, Brian Weatherson, and Stéphane Zuber. We also received useful feedback when the paper was presented, such as in the *Decisions, Ethics and Uncertainty Reading Group* at Hebrew University of Jerusalem (April 2018), the *Ethics and Uncertainty Workshop* at Hebrew University of Jerusalem (June 2018), the *Norms and Normativity Congress* at ENS Lyon (June 2018), and the yearly *Kick-Off Workshop* of Paris School of Economics (September 2018). Franz Dietrich acknowledges support by the French National Research Agency through the grant "Coping With Heterogeneous Opinions" and through an EUR grant. Brian Jabarian acknowledges support through a Global Excellence Gulbenkian Scholarship.

[2] The 'utility' in question must be something about which there is a fact and hence a possible uncertainty, such as rational desirability, personal interest, or true desire level (all of which are controversial concepts).

legal value, or artistic value. In this evaluation, one can face two fundamentally different types of uncertainty. Decision theory has so far focused on *empirical* uncertainty: uncertainty about empirical facts, states, or consequences of options. A doctor may be uncertain whether a particular treatment cures the patient. Philosophers have recently turned to so-called *normative* uncertainty: uncertainty about the correct standard of evaluation itself (e.g., Oddie 1994, Lockhart 2000, Jackson and Smith 2006, MacAskill 2014, Bradley and Drechsler 2014, Ross 2006, Sepielli 2006, 2009, Weatherson 2014, Grieves and Ord 2017, Lazar 2017, MacAskill and Ord forth.).[3] Even if the treatment is certain to cure the disease and known also in all its other empirical features – so that no empirical uncertainty whatsoever exists – the doctor can be uncertain about the moral value of the treatment, perhaps by wondering about the moral trade-off between the recovery and the (fully known) side effects of the treatment, or by hesitating between utilitarian and deontological evaluations.[4] Normative uncertainty can be more pressing and harder to resolve than empirical uncertainty – which makes it more surprising that formal decision theory so far neglects normative uncertainty. But is normative uncertainty truly distinct and meaningful?

Why ordinary choice theory does not yet capture normative uncertainty. Before proceeding, we briefly address two misunderstandings or prejudices one can have about the notion of normative uncertainty. Firstly, might normative uncertainty be reducible to empirical uncertainty after all, so that decision theorists could continue working with a single type of uncertainty? No doubt, many would suspect this. But a reduction fails, for conceptual and formal reasons. *Conceptually*, a reduction conflates fundamentally distinct phenomena. Uncertainty about value is not uncertainty about empirical features or consequences of options. Uncertainty whether a utilitarian or deontological valuation is correct is fundamentally evaluative uncertainty, compatible with full certainty about empirical features and consequences. *Formally*, can standard choice theory not already capture normative uncertainty, after suitably re-interpreting standard choice-theoretic models? It can not. By re-interpreting Savage's nature states as empirical-*normative* states, or re-interpreting von-Neumann-Morgenstern's lotteries as lotteries over empirical-*normative* outcomes, we change not just the interpretation, but also the formal game itself. Why? If nature states suddenly contain value information, then they anticipate the evaluation of consequences by encoding the utility of each consequence – something compatible with neither ordinary expected-utility theory nor even state-dependent expected-utility theory.[5] Analogously, classical expected-utility theory in von-Neumann-Morgenstern's lottery setting is inconsistent with building value information ('utility information') into outcomes, because classical utilities are free parameters, determined by preferences rather than pre-determined by information 'in' outcomes.

[3] Harsanyi's (1978) impartial observer can be interpreted as facing normative uncertainty.

[4] In general, the doctor may wonder (i) *which* properties of the treatment matter morally, and (ii) *how* they matter (Dietrich and List 2013, 2017).

[5] Choice theorists rarely allow utilities of consequences to depend on states, and where they do (i.e., in *state-dependent expected-utility theory*) they still take utilities to be up to the agent rather than written into states.

Is normative uncertainty meaningful? There can be uncertainty only where there are facts – but *are* there normative facts? Is there a 'right' standard of evaluation, an object of uncertainty? For the sake of argument, let 'value' be 'moral value' (but the issue is similar for other types of value). We cannot dig into ongoing meta-ethical debates here, but we hasten to stress that full-blown moral realism is not required to make sense of moral uncertainty. Moral uncertainty exists meaningfully under a range of meta-ethical positions. Moral facts could be subjective rather than objective: their existence could depend on attitudes of agents, be they ideal agents or even the decision-maker himself. Moral facts could be relative rather than universal, as claimed by cultural or social relativism. And moral facts could be understood in the sense of meta-ethical constructivisms. Error-theoretic positions should however be excluded, as they deny moral facts. It is debatable whether certain non-cognitivist (e.g., expressivist) positions are compatible with moral uncertainty, through some non-literal interpretation of 'moral uncertainty' which needs no moral facts (Weatherson 2014, Sepielli 2017).

2 Defining the problem and its expected-value solution

To set the stage in due precision, we now carefully define the framework in a standard version, and the expected-value approach. Later we shall add empirical uncertainty to the framework.

The objects of evaluation. We consider a non-empty set A of objects of evaluation, called '*options*'. They could be policy measures, social arrangements, income distributions etc. So far we leave open whether options contain empirical risk.

The competing valuations. Options have uncertain value, for instance because of competing normative intuitions or theories. Following the expected-value approach, our notion of value is numerical, not just comparative (for a defence, see MacAskill and Ord forth.). We thus represent a possible standard of evaluation by a function v, called a (first-order) *valuation* or *theory*, assigning to each option a in A a value $v(a)$ in $\mathbb{R}$. For instance a utilitarian valuation defines $v(a)$ as total happiness resulting from a. But not all functions from A to $\mathbb{R}$ are plausible valuations. Implausible valuations, like that in terms of resulting unhappiness, can be excluded from consideration. Let $\mathcal{V}$ be the set of valuations/theories considered, formally a finite non-empty set of functions from A to $\mathbb{R}$. $\mathcal{V}$ might contain just two valuations, a utilitarian and a Rawlsian one, or a deontological and an egalitarian one. Alternatively, $\mathcal{V}$ might consist of many valuations of similar type differing only in one parameter: prioritarian valuations with different degrees of prioritarianism, or egalitarian valuations with different degrees of inequality-aversion, or valuations of intertemporal (individual or social) well-being with different discounting of future well-being, or valuations (of risky options) with different degrees of risk-aversion, etc. If $\mathcal{V}$ is such a parametric family of valuations, the normative uncertainty is an uncertainty about the correct parameter value: the correct amount of prioritarianism, inequality-aversion, discounting, risk-aversion, etc.

Correctness probabilities of valuations. Each valuation v in $\mathcal{V}$ has a fixed correctness probability $Pr(v) \geq 0$, where $\sum_{v \in \mathcal{V}} Pr(v) = 1$. Probabilities capture degrees of belief about value (other interpretations are possible). The possibility to quantify normative uncertainty through sharp probabilities is a standard assumption.

Meta-valuations. What is the overall value of options given the normative uncertainty? Any answer is a *meta-valuation* or *meta-theory*, formally another function assigning to each option in A a value in $\mathbb{R}$. To distinguish it from first-order valuations, it is generically denoted as an upper-case V. So $V(a)$ represents a's overall value. Many potential meta-valuations come to mind. For instance, an option a's overall value might be its minimal value across all first-order theories of non-zero probability: $V(a) = \min_{v \in \mathcal{V}: Pr(v) \neq 0} v(a)$.[6]

Convention. We shall often say 'valuation' or 'theory' simpliciter, dropping 'first-order' or 'meta-', provided there is no ambiguity.

Expected value theory ('*EV*'). *This meta-theory evaluates each option $a \in A$ by its expected value:*

$$EV(a) = \sum_{v \in \mathcal{V}} Pr(v)v(a).$$

Here, overall value is average first-order value, weighted by correctness probabilities.

Measurability and comparability of value. The expected value theory is sometimes criticized for relying on numerical measurements and cross-theory comparisons of value. Measurability makes it meaningful for an option x to have value 7 under a valuation v ($v(x) = 7$), or be twice as valuable as another option y ($v(x) = 2v(y)$), or exceed y's value by 2 ($v(x) - v(z) = 2$), etc. Comparability makes it meaningful for option x to be equally valuable under two valuations v and v' ($v(x) = v'(x)$), or more valuable under v than option y is under v' ($v(x) > v'(y)$), etc.

Setting this debate aside, we show that the expected value theory is problematic *even if* value is fully measurable and cross-theory comparable. We thus retain this classic assumption throughout the paper (although it could be partly relaxed[7]). Note that comparability across valuations seems less problematic if $\mathcal{V}$ consists of valuations of similar type (e.g., egalitarian valuations with different degrees of inequality-aversion), but more debatable if $\mathcal{V}$ contains radically different valuations such as consequentialist and deontological ones.

[6] For convenience, we define meta-theories as value *functions*, not (binary) value *relations*, thereby invoking levels of, not just comparisons in, overall value. Everything could be restated using purely relational meta-theories.

[7] The expected value theory and its three alternatives introduced below do not need full measurability of value: value need only be measurable on an affine scale, so that valuations are only unique up to increasing affine transformation. Further, the expected value theory does not need full comparability of value across theories: unit comparability without level comparability suffices. Our impartial meta-theory however needs value comparability. Comparability and measurability are addressed by Bossert and Weymark (2004), and in the context of normative uncertainty by, e.g., Ross (2006) and Spirelli (2009).

Value versus von-Neumann-Mogenstern ('vNM') utility. The value of an option under a valuation in $\mathcal{V}$ should normally not be interpreted as a vNM utility (or expected vNM utility if the option is risky). To see why, consider a hedonistic theory v in $\mathcal{V}$. It evaluates riskless options by the amount of experienced pleasure generated by the option. For any $t \geq 0$, consider the option $a_t \in A$ of donating t coins to a charity. Let the agent's pleasure from donating be never saturated: the first donated coin gives as much pleasure as the second, the second as much pleasure as the third, etc. Then the value of donating t coins, $v(a_t)$, is linear in t. But the vNM utility of a_t may or not be linear in t, depending on the risk attitude of the evaluative theory in question: the vNM utility of a_t is linear, concave or convex in t (and in $v(a_t)$), depending on whether we consider risk-neutral, risk-averse or risk-prone hedonism, respectively.[8] Simply replacing v by a vNM utility function (or *expected*-vNM-utility function if A also contains risky options) is inappropriate, as we would then incorrectly measure pleasure gains: for instance, the vNM utility function of risk-averse hedonism rises less from a_1 to a_2 than from a_0 to a_1 (by concavity), although the pleasure gain is the same both times. The vNM approach is ill-suited for the measurement and inter-theory comparison of value. See Broome (1991), Weymark (1991) and Nissan-Rozen (2015) for various analyses of the controversial value-utility relationship.

3 Two problems of the expected value theory

This section raises two objections against expected value theory, and proposes some principles, interpretable either as general normative principles or as methodological principles for designing operational meta-theories. We draw on empirical uncertainty, besides normative uncertainty. The possibility that options carry empirical uncertainty – e.g., uncertainty about consequences – is often acknowledged and explicitly allowed in the literature (Weatherson 2014, Nissan-Rozen 2016, MacAskill and Ord forth.), although empirical uncertainty is usually not formalized (an exception is Bradley and Drechsler 2013). Our objections and principles will so far be stated informally, as we postpone the formalization of empirical uncertainty to Section 5.

3.1 Overruling beliefs about the correct risk attitude

Consider two options a and b in A. Think of them as containing no empirical risk: their features are fully known. Let there be just two competing valuations v and v' in $\mathcal{V}$, each of correctness probability $\frac{1}{2}$. Table 1 shows how each option is evaluated by v and v', and by the expected value (meta-) theory.

We can make the example concrete. Ann, Bob, and Claire suffer from a disease. Ann owns 2g of a medicine, which is just enough to cure her. Bob and Claire would each need just 1g of that medicine to be cured. The agent (e.g., a public health authority) can either not intervene, so that Ann is cured by her medicine while Bob and Claire

[8] *Risk-neutral, -averse, and -prone hedonism* is an extension of hedonism to risky prospects which regards the value of a risky prospect as being equal, below, or above the expected amount of resulting hedonic pleasure, respectively. This can be stated formally if A contains risky options of the sort 'donating *so-and-so* many coins with *such-and-such* probabilities'.

	evaluation by		
option	v	v'	meta-theory EV
a	2	2	$2 = \frac{1}{2}2 + \frac{1}{2}2$
b	4	0	$2 = \frac{1}{2}4 + \frac{1}{2}0$

Table 1: Same overall evaluation despite different levels of value risk

stay ill. This is option a. Or the agent confiscates Ann's medicine and redistribute it among Bob and Claire, so that Bob and Claire get cured while Ann stays ill. This is option b. Curing someone contributes two units of well-being to that person. Let v be a utilitarian theory which evaluates options by total resulting well-being. So option a has value 2 (one person cured) while b has value 4 (two persons cured). Theory v' is a deontological theory which also attaches importance to the respect of property. It evaluates options by total resulting well-being, *minus* 4 in case of property violation. So option a has value 2 (one person cured, property respected), while b has value $0 = 4-4$ (two persons cured, property violated).

The options a and b have the same expected value of 2. But assigning same overall value to a and b is problematic, as b contains normative uncertainty while a does not. By giving b overall value 2, one is neutral to the normative risk in b. Such meta-theoretic risk-neutrality can overrule a unanimous risk attitude among first-order valuations. It can indeed happen that

- some option c displays exactly the same risk as b in terms of resulting value, except that the source of risk is empirical rather than normative,
- the first-order valuations v and v' evaluate c identically, at a same risk premium,
- although the expected value (meta-) theory evaluates b at no risk premium.

For example, assume both valuations v and v' are risk-averse: risky options – options which could result in different empirical worlds according to certain probabilities – are evaluated below the expected value of the resulting world. Let risky option c result in a 'positive' or a 'negative' world, each with probability $\frac{1}{2}$. These worlds have value 4 and 0 respectively, according to both v and v'.[9]

Option c can be made concrete. Imagine a second medicine which can only cure Bob, for generic reasons say. With probability $\frac{1}{2}$, that medicine has a terrible side effect reducing Bob's well-being by 4 units. Option c consists in giving Bob that medicine, without redistributing Ann's medicine. c results

- either in the 'positive' world without side effect, of value 4 under both valuations (two persons cured, no side effect, property respected),
- or in the 'negative' world with side effect, of value $0 = 4-4$ under both valuations (two persons cured, one suffering side effect, property respected).

Being risk-averse, v and v' evaluate c below the expected resulting value of $2 = \frac{1}{2}4 + \frac{1}{2}0$. We assume $v(c) = v(c') = 1$, amounting to a risk premium of $2 - 1 = 1$.

The options b and c both lead to the same *value prospect*: the prospect that the resulting value is 4 with probability $\frac{1}{2}$ and 0 with probability $\frac{1}{2}$. So b and c display the

[9] So a hypothetical option which surely results in the 'positive' world has value 4, while a hypothetical option which surely results in the 'negative' world has value 0, under both theories.

same risk in terms of resulting value, although the source of risk is normative for b and empirical for c. Since this risky value prospect justifies a risk premium of 1 according to v and v' (given how v and v' evaluate c), one would have expected the meta-theory to adopt this unanimous risk aversion, even where the source of risk is normative. This suggests evaluating b at a risk premium – against the expected value theory.

In sum, the expected value theory can create the awkward situation of neutrality to normative risk paired with aversion to empirical risk. Such a hybrid risk attitude is at least question-begging, as one wonders what would justify neutrality to normative risk if one should certainly be averse to empirical risk.

More systematically speaking, the expected-value theory violates the following principle:

Risk-Attitudinal Unanimity Principle (stated informally): *If there is certainty about the correct risk-attitude, i.e., all valuations in $\mathcal{V}$ of positive probability have same risk attitude, then the meta-theory adopts this risk attitude (even towards normative risk).*

The following broader principle also covers cases of risk-attitudinal heterogeneity or uncertainty:

Risk-Attitudinal Impartiality Principle (stated informally): *The meta-theoretic risk attitude reflects impartially the beliefs about the correct risk-attitude captured by the correctness probabilities of first-order theories.*

This principle requires forming a compromise between the risk attitudes of the first-order theories. The more likely a risk attitude is to be correct, i.e., the higher the total probability of first-order theories with that risk attitude is, the more weight that risk attitude should get in the meta-theory. Although details are postponed to Section 6, it should already be clear that, under any plausible interpretation, the principle implies the Risk-Attitudinal Unanimity Principle – because a certainty that a particular risk attitude is correct is 'respected impartially' only if that risk attitude is adopted.

3.2 Relying on information outside the value prospect

The expected value theory also violates a very basic idea, captured by the following principle.

Value-Prospect Principle (stated informally): *An option's overall value is determined by its value prospect, i.e., its (probability distribution of) possible values after resolution of uncertainty.*

Before motivating this principle, let us see how the expected value theory violates it. The options b and c of our leading example have the same value prospect denoted $4_{50\%}0_{50\%}$: the resulting value is 4 with probability $\frac{1}{2}$ and 0 with probability $\frac{1}{2}$ – for b because of normative uncertainty about the value of the (single possible) outcome,

and for c because of empirical uncertainty about which outcome obtains (each outcome having uncontroversial value).

	value	evaluation of option by		
option	prospect	v	v'	meta-theory EV
b	$4_{50\%}0_{50\%}$	4	0	$2 = \frac{1}{2}4 + \frac{1}{2}0$
c	$4_{50\%}0_{50\%}$	1	1	$1 = \frac{1}{2}1 + \frac{1}{2}1$

Table 2: Different overall evaluation despite same value prospect

Being risk-averse, v and v' evaluate c at 1, below c's expected outcome value of $2 = \frac{1}{2}4 + \frac{1}{2}0$. This leads to a lower overall value for c than for b – against the Value-Prospect Principle.

The principle is plausible because value prospects seem to capture everything relevant: they represent our expectations about the value of the resulting world, where 'worlds' capture all normatively relevant features and are evaluated in an all-things-considered way by each valuation in $\mathcal{V}$. From a consequentialist perspective, worlds are consequences, and nothing but the value of consequences matters. Consequentialism 'almost' implies our principle – 'almost', because it extends consequentialism to risky cases. That natural extension takes consequentialism to require that options be evaluated solely based on the probability distribution of the (value of the) consequence – which *is* our principle. But even outside the consequentialist paradigm our principle is plausible. The fact that worlds go beyond consequences does not change the fact that worlds contain all normatively relevant features and are being comprehensively evaluated by each valuation in $\mathcal{V}$ – so that it remains plausible that two options have same overall value if they have same value prospect, i.e., are indistinguishable in the probabilities with which final values are achieved.[10]

4 Going non-expectational is no solution

The expected value theory needs revision, as it violates the principles of which at least one – the Risk-Attitudinal Unanimity Principle – seems incontestable. *How* should one revise it? As we have complained, it is neutral to normative risk, rather than reflecting the attitudes of the first-order theories to empirical risk. Surely the first attempt to fix this problem is to give up the expectational (i.e., risk-neutral) form: to aggregate the option values $v(a)$ ($v \in \mathcal{V}$) in some other way which purportedly reflects the first-order risk attitudes. The new aggregate option value should be *below* the expected option value if all first-order valuations (of non-zero correctness probability) are risk-averse, and *above* the expected option value if all first-order valuations (of non-zero correctness probability) are risk-prone.

Surprisingly, this natural approach fails. Consider again our lead example, with its risk-averse valuations v and v'. According to this approach, the overall value of an option o is not its expected value $\frac{1}{2}v(o) + \frac{1}{2}v'(o)$, but some *non-expectational* aggregate

[10] This defence of the principle relies on the classic assumption that value is comparable across theories in $\mathcal{V}$, as value prospects 'mix' across different theories in $\mathcal{V}$.

$F(v(o), v'(o))$ of $v(o)$ and $v'(o)$. Here the aggregation functional F maps any combination $(k, k') \in \mathbb{R} \times \mathbb{R}$ of first-order values to an overall value $F(k, k')$, where to respect risk-aversion $F(k, k') < \frac{1}{2}k + \frac{1}{2}k'$ (unless $k = k'$, the case of certainty about the option value). The amount by which $F(k, k')$ falls short of the average $\frac{1}{2}k + \frac{1}{2}k'$ represents a risk premium. For instance, consider the option b of redistributing the medicine among Bob and Claire. Its value is 4 under v and 0 under v'; so here $(k, k') = (4, 0)$. The overall value of b would thus be $F(v(b), v'(b)) = F(4, 0) < 2$.

What is the problem? Consider another option d which (unlike b) contains empirical uncertainty: it has two possible outcomes of probability 50% each, of values 7 and 3 under v and of values 3 and -1 under v'. Following the risk-aversion of v and v', let $v(d)$

	value prospect of option			evaluation of option by			
option	under v	under v'	overall	v	v'	EV	NEV
b	$4_{100\%}$	$0_{100\%}$	$4_{50\%}0_{50\%}$	4	0	2	$F(4,0)$
d	$7_{50\%}3_{50\%}$	$3_{50\%}(-1)_{50\%}$	$7_{25\%}3_{50\%}(-1)_{25\%}$	4	0	2	$F(4,0)$

Table 3: A non-expectational value theory ('NEV') cannot distinguish between the options b and d despite their different overall value prospects.

be 4 (below the expected value $\frac{1}{2}7 + \frac{1}{2}3 = 5$) and let $v'(d)$ be 0 (below the expected value $\frac{1}{2}3 + \frac{1}{2}(-1) = 1$). Option d is indistinguishable from b in terms of first-order evaluations: $v(d) = v(b)$ and $v'(d) = v'(b)$. This forces d to have the same overall value as b, namely again $F(4, 0)$ (see Table 3). However we see no compelling argument for overall indifference between d and b. Option d has a more 'disparate' value prospect than b, namely $7_{25\%}3_{50\%}(-1)_{25\%}$ instead of $4_{50\%}0_{50\%}$.[11] In result, d is more risky than b under many risk measures.[12] Since risk should influence overall value, d may have to be evaluated differently from b.

The lesson is this: the problem of the expected value theory is not so much *how* it aggregates the first-order option values (namely expectationally), but more fundamentally *that* it builds on first-order option values. To fix the theory, we must 'unpack' options and dig into their empirical risk structure. Let us do this.

5 Three new meta-theories

We now introduce new meta-theories. One of them – the *impartial value theory* – will satisfy all our principles, suitably interpreted. Before stating the meta-theories, we enrich the framework by empirical uncertainty.

[11] For instance, the probability of value 7 is the probability of the 'better' outcome (50%) times the correctness probability of the valuation v (50%).

[12] If the risk in an option is measured by the variance (second moment) of the value prospect, then b counts as less risky than d. Indeed, b's value prospect $4_{50\%}0_{50\%}$ has variance $\frac{1}{2}(4-2)^2 + \frac{1}{2}(0-2)^2 = 4$, while d's value prospect $7_{25\%}3_{50\%}(-1)_{25\%}$ has variance $\frac{1}{4}(7-3)^2 + \frac{1}{2}(3-3)^2 + \frac{1}{4}((-1)-3)^2 = 8$. More generally, d counts as more risky than b if we measure risk by the m^{th} (absolute) moment of the value prospect for any order $m \in (1, \infty]$. If by contrast risk is measured by the first (absolute) moment of the value prospect, then b and d count as equally risky, since b's value prospect has first moment $\frac{1}{2}|4-2| + \frac{1}{2}|0-2| = 2$ and d's has first moment $\frac{1}{4}|7-3| + \frac{1}{4}|(-1)-3)| + \frac{1}{2}|3-3| = 2$.

5.1 Formalising empirical uncertainty: options as lotteries

Although empirical uncertainty is usually not formalized, the problem of choice under normative uncertainty and its expected-value solution are explicitly meant to cover options carrying empirical uncertainty (Nissan-Rozen 2015, Williams 2017, MacAskill and Ord forth.). Empirical uncertainty can in fact be easily formalized, and this within the existing framework of normative uncertainty. We simply assume that options in A are not primitive objects, but lotteries over a fixed set X of 'worlds'. Formally, each option a in A is a function assigning to each world x in X a probability $a(x)$ in $[0, 1]$ such that $\sum_{x \in X} a(x) = 1$, where (for simplicity) only finitely many worlds x have non-zero probability $a(x)$. An option is *riskless* if some world has probability one, and *risky* otherwise.

A world represents a possible empirical state of affairs after empirical uncertainty is resolved. A world need not contain everything: it need neither contain empirical facts that are normatively irrelevant (under the theories in $\mathcal{V}$), nor of course *normative* facts (about the value of worlds).

In one application, worlds are consequences of options after resolution of empirical uncertainty. This restricts our model to the case where all valuations in $\mathcal{V}$ are consequentialist. Did we only have consequentialist valuations in mind, we could say 'outcome' for 'world'. Alternatively, worlds could also contain non-consequence facts, such as facts about intentions, the choice context, or whether the chosen option involves 'acting' or 'omitting'. This opens the model to non-consequentialist valuations in $\mathcal{V}$.[13]

A need not contain *all* lotteries over (finitely many) worlds.[14] All we assume is that A contains for each world in X a corresponding riskless option which yields that world. We can thus take each valuation v to evaluate not just options, but also worlds: the value of a world x is the value of the corresponding riskless option a, formally $v(x) = v(a)$.

Ever since von Neumann and Morgenstern (1944) it is common to evaluate lotteries by the expectation of some underlying 'utility' function over worlds. Valuations in $\mathcal{V}$ may *but need not* take such an expected-utility form, as already mentioned informally in Section 1.[15]

[13] Normative uncertainty for non-consequentialist theories is addressed by Barry and Tomlin (2016) and Tenenbaum (2017).

[14] The set of options A *could* be so 'rich' as to contain all lotteries over (finitely many) worlds, in line with common von-Neumann-Morgenstern-type lottery setups. But if worlds go beyond consequences – which they must if some normative theories in V are non-consequentialist – then certain lotteries over worlds may become meaningless as options and should then be excluded from A. If worlds for instance contain context-related information, then a meaningful option can yield only context-wise identical worlds (a choice happens in just one context), so that many lotteries are excluded from A. Similarly, if worlds contain information about the chooser's intention, then a meaningful choice option can yield only intention-wise identical worlds (one cannot have two intentions simultaneously), which again limits the set of options A.

[15] So, a v in $\mathcal{V}$ need not evaluate options by the expectation of any (von-Neumann-Morgenstern) 'utility' function of worlds, i.e., there need not exist a function u on X such that $v(a) = \sum_{x \in X} a(x)u(x)$ for all options (lotteries) a in A. For one, the order over lotteries induced by v need not satisfy von-Neumann-Morgenstern's axioms, or equivalently, need not be representable by an expected-utility function. For another, even if that order is representable by an expected-utility function (unique up to increasing linear transformation), then v need not be such an expected-utility function, despite

5.2 Value prospects defined

A 'value prospect' is a prospect of achieving certain value levels with certain probabilities, for instance achieving value 4 with probability 1/2 and value 0 with probability 1/2. Formally, a *value prospect* is a lottery over value levels, i.e., a function p assigning to each value k in $\mathbb{R}$ a probability $p(k)$ in $[0,1]$ such that $\sum_{k\in\mathbb{R}} p(k) = 1$, where (for simplicity) only finitely many values k have non-zero probability $p(k)$.

Each option generates a value prospect. The final value resulting from an option a depends on the resulting world (empirical uncertainty) and the correct valuation (normative uncertainty). So it depends on the world/valuation combination (x, v) in $X \times \mathcal{V}$. The probability of achieving value 7 is the sum-total probability of all world/value combinations (x, v) such that $v(x) = 7$. Here the probability of each combination (x, v) is the probability $a(x)$ of the world *times* the correctness probability $Pr(v)$ of the valuation. Formally: the **value prospect of an option** $a \in A$ is the value prospect p_a under which any value $k \in \mathbb{R}$ has probability

$$p_a(k) = \text{ 'probability of a world of value } k\text{' } = \sum_{(x,v)\in X\times\mathcal{V}: v(x)=k} \underbrace{a(x)Pr(v)}_{\text{prob. of } (x,v)}.$$

This definition can be restricted by eliminating either normative or empirical uncertainty:

- To eliminate normative uncertainty, we fix a valuation in $\mathcal{V}$ and average only across worlds. Formally: the **value prospect of an option** $a \in A$ **given a valuation** $v \in \mathcal{V}$ is the value prospect $p_{a,v}$ under which any value $k \in \mathbb{R}$ has probability

  $$p_{a,v}(k) = \text{ 'probability of a world of value } k \text{ under valuation } v\text{' } = \sum_{x\in X: v(x)=k} a(x).$$

- To eliminate empirical uncertainty, we fix a world and average only across valuations in $\mathcal{V}$. This effectively yields the value prospect *of a world*, not an option. Formally: the **value prospect of a world** $x \in X$ is the value prospect p_x under which any value $k \in \mathbb{R}$ has probability:[16]

  $$p_x(k) = \text{ 'probability that } x \text{ has value } k\text{' } = \sum_{v\in\mathcal{V}: v(x)=k} Pr(v).$$

In sum, each option a yields certain values with certain probabilities, as described generally by p_a, and more restrictively by $p_{a,v}$ or p_x when conditionalizing on a particular valuation v or world x, respectively.

5.3 The new meta-theories illustrated informally

Consider our leading example, with its two risk-averse valuations v and v' of correctness probability $\frac{1}{2}$ each. The riskless option b of redistributing the medicine to Bob and Claire yields a sure world x; it is denoted $x_{100\%}$. As its value is 4 under v and 0 under

representing the same order and hence being ordinally equivalent to an expected-utility function..

[16] p_x equals the value prospect p_a of the risk-free option a generating x.

v'; its value prospect is denoted $4_{50\%}0_{50\%}$. Its overall value according to expected value theory is $\frac{1}{2}4 + \frac{1}{2}0 = 2$ – something we have criticized for being neutral to normative risk although v and v' are risk-averse. Attempting to repair expected value theory by aggregating the possible values 4 and 0 'non-expectationally' – into some overall value below 2 – is a non-starter by Section 4. We should rather aggregate other information than first-order option values. *What* other information? There are three salient approaches. To illustrate them, consider the risky option c of curing Bob with a different medicine that has either no side effect (the 'positive' world y, of probability 50%) or a severe side effect (the 'negative' world z, of probability 50%); formally, $c = y_{50\%}z_{50\%}$. Both v and v' assign value 4 to y and value 0 to z; so c's value prospect is $4_{50\%}0_{50\%}$. Being risk-averse, v and v' both evaluate c at 1, below the expected value of 2. Although both options b and c have same value prospect, the source of uncertainty is normative for b and empirical for c. Table 4 shows how b and c are evaluated by four

	value prospect of option			evaluation of option by					
option	under v	under v'	overall	v	v'	EV	FEV	IV	DEV
$b = x_{100\%}$	$4_{100\%}$	$0_{100\%}$	$4_{50\%}0_{50\%}$	4	0	2	2	1	1
$c = y_{50\%}z_{50\%}$	$4_{50\%}0_{50\%}$	$4_{50\%}0_{50\%}$	$4_{50\%}0_{50\%}$	1	1	1	2	1	2

Table 4: The four meta-theories in the medical example

meta-valuations, namely the expected value theory EV and our three alternative value theories, the *fully expectational*, *impartial*, and *dual expected* value theories, denoted FEV, IV, and DEV, respectively. All four meta-theories define overall value as the expected value of *some* object. That object – called the 'focus of evaluation' – differs:

EV: Here the focus of evaluation is the option itself. So the overall value of an option is the average value of that option itself, i.e., $\frac{1}{2}4 + \frac{1}{2}0 = 2$ or $\frac{1}{2}1 + \frac{1}{2}1 = 1$, respectively.

FEV: Here the focus of evaluation is the world, i.e., the state of affairs after resolution of empirical uncertainty. Computing the average value of the world requires averaging not just across valuations in $\mathcal{V}$ (normative uncertainty), but also across worlds (empirical uncertainty). It may turn out – and *does* turn out for our two options – that only one dimension of averaging is needed, as only one source of uncertainty exists. Option b involves just normative uncertainty: it surely yields world x, of value 4 or 0. Option c involves just empirical uncertainty: it yields world y of sure value 4 or world z of sure value 0. For both options, the average value is $\frac{1}{2}4 + \frac{1}{2}0 = 2$.

IV: Here the focus of evaluation is the value prospect of the option, which is $4_{50\%}0_{50\%}$ both times. So we must calculate how the value prospect $4_{50\%}0_{50\%}$ is evaluated on average by v and v'. But first, how does a valuation (v or v') evaluate value prospects rather than options? Value prospects are evaluated like their corresponding options: v takes $4_{50\%}0_{50\%}$ to be the value prospect of option $y_{50\%}z_{50\%}$, so that $v(4_{50\%}0_{50\%})$ reduces to $v(y_{50\%}z_{50\%}) = 1$; and v' also takes $4_{50\%}0_{50\%}$ to be the value prospect of $y_{50\%}z_{50\%}$, so that $v'(4_{50\%}0_{50\%})$ reduces to $v'(y_{50\%}z_{50\%}) = 1$. So the average value of $4_{50\%}0_{50\%}$ is $\frac{1}{2}v(4_{50\%}0_{50\%}) + \frac{1}{2}v'(4_{50\%}0_{50\%}) = \frac{1}{2}1 + \frac{1}{2}1 = 1$.

DEV: Here the focus of evaluation is the value prospect of the world, not of the option.

For each option, we must calculate how the world's value prospect is evaluated on average. Like for FEV, this requires averaging across both worlds and valuations – though for our options b and c one dimension of averaging drops out, as b has no empirical uncertainty while c has no normative uncertainty about evaluating value prospects of relevant worlds. Concretely, the option $b = x_{100\%}$ surely yields world x, whose value prospect $4_{50\%}0_{50\%}$ is evaluated at 1 by both (risk-averse) valuations v and v', as seen earlier. So the average value is $\frac{1}{2}1 + \frac{1}{2}1 = 1$. The option $c = y_{50\%}z_{50\%}$ *either* yields world y, whose value prospect $4_{100\%}$ has value 4 under both v and v'; *or* yields world z, whose value prospect $0_{100\%}$ has value 0 under both v and v'. So the average value is $\frac{1}{2}4 + \frac{1}{2}0 = 2$.

The risk-attitudinal rationale of each meta-theory. Each meta-theory has a focus of evaluation that is tailor-made for a particular risk attitude:

	normatively risk-neutral	normatively risk-averse
empirically risk-neutral	FEV	DEV
empirically risk-averse	EV	IV

Table 5: The risk attitudes of the four meta-theories in our example with risk-averse first-order theories

EV: Here the focus of evaluation – the option – captures empirical risk by being a lottery over worlds, but captures no normative risk. Through applying the (risk-averse) valuations in $\mathcal{V}$ to options, EV discounts for empirical risk, not for normative risk: it is averse to empirical risk only. Options without empirical risk like b are evaluated without discount, which explains why b gets higher value than c.

FEV: Here the focus of evaluation – the world – captures neither empirical, no normative risk. Being riskless, worlds are evaluated without risk premium by the valuations in $\mathcal{V}$. Hence FEV contains no risk premia: it is globally risk-neutral. This explains why both options b and c get the 'high' value of 2.

IV: Here the focus of evaluation – the value prospect of the option – captures both empirical and normative risk. By applying the (risk-averse) valuations in $\mathcal{V}$ to the value prospect, FEV discounts for both types of risk, hence is globally risk-averse. This explains why both b (involving normative risk) and c (involving empirical risk) get the 'low' value of 1.

DEV: Here the focus of evaluation – the value prospect of the world – captures normative, but not empirical risk. By applying the (risk-averse) valuations in $\mathcal{V}$ to value prospects of worlds, DEV discounts for normative, but not empirical risk, hence is averse to normative risk only. This explains why b (involving normative risk) gets a higher value than c (involving empirical risk).

5.4 The new meta-theories defined formally

We start with the fully expectational value theory. It is not just 'normatively expectational' (like expected value theory EV), but also 'empirically expectational'. It averages

across worlds as well as valuations, hence across world/valuation combinations (x, v) in $X \times \mathcal{V}$. Each combination (x, v) yields a value, $v(x)$. It is weighted by the probability of the combination, i.e., the probability $a(x)$ of world x *times* the probability $Pr(v)$ of valuation v. Formally:

Fully expectational value theory ('FEV'). *This meta-theory evaluates each option $a \in A$ by the expected value of the resulting world:*

$$FEV(a) = \sum_{(x,v) \in X \times \mathcal{V}} \underbrace{a(x)Pr(v)}_{prob.\ of\ (x,v)} v(x) \text{ (the 'fully expectational value' of } a).$$

The impartial value theory operates at a different level: not the level of options (EV) or worlds (FEV), but the level of value prospects. This is possible because valuations v in $\mathcal{V}$ can be taken to evaluate not just options (and hence worlds, i.e., riskless options), but also value prospects. The value of a value prospect is simply the value of options having that value prospect. This definition however presupposes that (i) there *exists* an option in A with that value prospect under v, and (ii) any two such options are evaluated equally by v. Condition (i) essentially means that there are sufficiently many options in A, a typical decision-theoretic richness assumption. Condition (ii) holds for most or all natural first-order theories; it is the analogue of our Value-Prospect Principle, for first-order theories rather than meta-theories. We now formally state the background assumption (i)-(ii), followed by the definition which it enables:

Assumption: Hereafter, for each valuation v in $\mathcal{V}$ and value prospect p, there is a corresponding option a in A whose value prospect $p_{a,v}$ is p, and any two such options a have same value $v(a)$.

Definition 1 *Under a valuation v in $\mathcal{V}$, the **value of a value prospect** p – denoted $v(p)$ – is the value $v(a)$ of options $a \in A$ whose value prospect under v is p, i.e., $p_{a,v} = p$.*

The 'impartial value' is the average value of the value prospect:

Impartial value theory ('IV'). *This meta-theory evaluates each option $a \in A$ by the expected evaluation of its value prospect:*

$$IV(a) = \sum_{v \in \mathcal{V}} Pr(v)v(p_a) \text{ (the 'impartial value' of } a).$$

The dual expected value theory has yet another focus of evaluation, which is neither the option (EV), nor the world (FEV), nor the value prospect of the option (IV), but the value prospect of the world. We thus form the average evaluation of the value prospect p_x, across worlds x and valuations v, i.e., across world/valuation combinations (x, v). In this weighted average, each combination (x, v) is weighted by its probability, the product of the probabilities $a(x)$ of x and $Pr(v)$ of v:

Dual expected value theory ('*DEV*'). *This meta-theory evaluates each option $a \in A$ by the expected value of the value prospect of the world:*

$$DEV(a) = \sum_{(x,v) \in X \times \mathcal{V}} \underbrace{a(x)Pr(v)}_{prob.\ of\ (x,v)} v(p_x) \text{ (the dual expected value of } a).$$

The distinctive mark of DEV is that it handles risk in the opposite way from EV. It applies valuations in $\mathcal{V}$ to world-value-prospects p_x ('normative lotteries') rather than to the option a (an 'empirical lottery'), thereby delegating the normative-risk attitude rather than the empirical-risk attitude to the first-order theories. Sections 6 and 7 will flesh this out formally.

5.5 The four meta-theories as ex-ante or ex-post theories

The four meta-theories are interpretable in terms of the famous ex-ante and ex-post approaches in ethics and aggregation theory under uncertainty.[17] These approaches refer to whether competing evaluations of uncertain prospects (e.g., evaluations by different individuals) are aggregated before or after resolution of uncertainty. We work with two types of uncertainty – empirical and normative – and hence have four possible approaches, depending on whether we go ex-ante or ex-post on each type:

	normatively ex-post	normatively ex-ante
empirically ex-post	FEV	DEV
empirically ex-ante	EV	IV

Table 6: Interpreting the meta-theories as ex-ante or ex-post theories

FEV: The fully expectational value theory is a fully ex-post theory, as it aggregates evaluations of worlds x, i.e., riskless states of affairs.

IV: The impartial value theory is a fully ex-ante theory, as it aggregates evaluations of the value prospect p_a, a state of unresolved empirical and normative uncertainty.

EV: The expected value theory is an empirically ex-ante theory, as it aggregates evaluations of the option a, a state of unresolved empirical uncertainty.

DEV: The dual expected value theory is a normatively ex-ante theory, as it aggregates evaluations of world-value-prospects p_x, i.e., states of unresolved normative uncertainty.

Given the importance of the ex-ante and ex-post approaches in ethics, all four meta-theories are obvious 'candidate theories' which deserve our attention.

5.6 In defence of the expectational form of IV

The impartial value theory IV has something in common with the classic expected value theory EV: the expectational or linear form. Indeed, IV builds a linear average

[17] There is an on-going debate about which approach is more appropriate. For instance, Diamond's (1976) famous objection against Harsanyi-type axiomatic utilitarianism assumes that ex-ante equality matters, something put into question by others (e.g., McCarthy 2006, 2008). The ex-ante and ex-post approaches to social well-being are compared by, e.g., Fleurbaey (2010), Fleurbaey and Voorhoeve (2016), and Fleurbaey and Zuber (2017).

of the $v(p_a)$'s ($v \in \mathcal{V}$), not any geometric average or other non-linear compromise. While EV's linearity causes the questionable neutrality to normative risk, IV's linearity does not cause any neutrality to normative or empirical risk. IV globally respects risk-attitudinal beliefs, as formally established later.

Aggregating the $v(p_a)$'s in some non-linear way – in an attempt to (even better) respect risk-attitudinal beliefs – could have the converse effect of 'overshooting' the risk-attitude, because of double-risk-discounting. Why? Assume all $v \in \mathcal{V}$ are risk-averse, and suppose in response the $v(p_a)$'s ($v \in \mathcal{V}$) were aggregated sub-linearly, into some overall value $IV^*(a) < IV(a)$. The meta-theory IV^* can be more risk-averse than all valuations $v \in \mathcal{V}$. Each value $v(p_a)$ ($v \in \mathcal{V}$) already contains a risk premium for all risk in a, empirical and normative, and hence so does $IV(a)$. Reducing $IV(a)$ further to $IV^*(a)$ imposes another risk premium – a second one. One thereby becomes more risk-averse at the meta-level than is certainly correct at the first-order level.[18]

This said, we do not categorically insist on linearity. We insist on aggregating the $v(p_a)$'s ($v \in \mathcal{V}$) rather than the $v(a)$'s ($v \in \mathcal{V}$), but we only propose a weighted linear average as the most natural approach. Other approaches might be defensible, if they somehow avoid double-risk-discounting. One might more explicitly call IV the '*linear* impartial value theory', which falls into the class of '*generalized* impartial value theories', i.e., of meta-theories IV^* which define the overall value of options a by aggregating the $v(p_a)$'s in *some* (possibly non-linear) way.[19]

6 The impartial value theory addresses both objections

This section and the next formally substantiate previous arguments and claims. The present section shows how the impartial value theory avoids both problems of the expected value theory raised in Section 3, after formally re-stating the principles proposed in Section 3.

6.1 Which meta-theories respect risk-attitudinal beliefs?

We now show that the impartial value theory IV satisfies Section 3's two risk-attitudinal principles, while EV, FEV and DEV do not. Only IV forms a genuine compromise between the risk attitudes of the competing valuations, thereby respecting risk-attitudinal beliefs. The fully expectational value theory FEV blatantly violates risk-attitudinal beliefs, by being globally risk-neutral; and the ordinary and dual expected value theories EV and DEV respect risk-attitudinal beliefs w.r.t. just empirical risk (EV) or just normative risk (DEV).

But first, what are 'risk-attitudes' of normative theories? Risk analysis has a long tradition in decision theory, where the bearers of risk-attitudes are usually individuals, not normative theories (e.g., Weirich 1986, 2004 Buchak 2013). Risk attitudes are often

[18] By contrast, correcting the classic expected value $EV(a)$ by subtracting a risk premium need not lead to double-risk-discounting, because $EV(a)$ does not yet contain any premium for normative risk. Yet a non-linear aggregation of the $v(a)$'s has different problems, explained in Section 4.

[19] Formally, $IV^*(a) = F(v(p_a))_{v \in \mathcal{V}})$ for all $a \in A$, for some fixed aggregation function F from $\mathbb{R}^{\mathcal{V}}$ to $\mathbb{R}$ (on which one might impose regularity conditions, such as increasingness).

defined through comparing the evaluation of risky objects with certain expectational evaluations. We adapt this approach to the realm of ethics and normative uncertainty. So we compare the value $v(a)$ of a given option a under a given theory v with the expected value of the resulting world, $\sum_{x \in X} a(x)v(x)$. Depending on whether $v(a)$ matches, exceeds, or falls below that expectation, the evaluation of a is risk-neutral, -prone, or -averse. Formally:[20]

Definition 2 *A valuation v in $\mathcal{V}$ is **risk-neutral (-averse, -prone) towards option** $a \in A$ if it evaluates a at (below, above) the expected value across worlds, i.e.,*

$$v(a) = (<, >) \sum_{x \in X} a(x)v(x).$$

*The **risk premium for** a or **degree of risk-aversion towards** a is the amount $prem_v(a)$ by which a's value falls below a's expected resulting value:*

$$prem_v(a) = \sum_{x \in X} a(x)v(x) - v(a).$$

A risk premium is a value discount due to risk. Its sign indicates whether there is risk aversion, neutrality, or proneness. Risk attitudes of *meta*-theories can be defined similarly, except that we must incorporate normative uncertainty by forming the expectation not just over worlds (empirical uncertainty), but also over valuations (normative uncertainty). Formally:

Definition 3 *A meta-valuation V is **risk-neutral (-averse, -prone) towards an option** $a \in A$ if it evaluates a at (below, above) the expected value across worlds and valuations, i.e.,*

$$V(a) = (<, >)\ FEV(a) = \sum_{(x,v) \in X \times \mathcal{V}} \underbrace{a(x)Pr(v)}_{prob.\ \ of\ (x,v)} v(x).$$

*The **risk premium for** a or **degree of risk aversion towards** a is the amount by which a's value falls below a's fully expectational value:*

$$Prem_V(a) = FEV(a) - V(a).$$

In principle, a valuation or meta-valuation could be risk-averse towards some options and risk-neutral or -prone towards others. If such jumps are absent, we can talk of risk aversion, neutrality or proneness *simpliciter*:

[20] This definition owes its plausibility to the assumption that theories in $\mathcal{V}$ measure value on an absolute scale. Accordingly, comparing value differences is meaningful: a rise in value from 0 to 1 represents the same gain as one from 100 to 101. By contrast, a von-Neumann-Morgenstern function does not measure value on an absolute scale. There is intuitive compatibility between risk-*aversion* and evaluating lotteries by the *expectation* of a von-Neumann-Morgenstern function. (Indeed, economists often interpret agents as risk-averse if they evaluate money lotteries by the expectation of a concave von-Neumann-Morgenstern function.)

Definition 4 *A theory v in $\mathcal{V}$ or meta-theory V is **risk-neutral (-averse, -prone)** if it is risk-neutral (-averse, -prone) towards all options with risky value prospect.*[21]

Remark 1 *The fully expectational value theory FEV is risk-neutral.*

Just as we can apply valuations in $\mathcal{V}$ to value prospects rather than options (Section 5.4), so we can apply risk premia to value prospects rather than options:

Definition 5 *Given a valuation v in $\mathcal{V}$, the **risk premium for a value prospect** p or **degree of risk aversion towards** p, denoted $prem_v(p)$, is the risk premium for options a with value prospect $p_{a,v} = p$.*

Remark 2 *$prem_v(p)$ can be expressed as the amount by which p's value falls below p's expectation:*

$$prem_v(p) = Exp(p) - v(p) = \sum_{k \in \mathbb{R}} p(k)k - v(p).$$

We now formally state the Risk-Attitudinal Unanimity Principle of Section 3. We state the principle in two versions, depending on whether we take 'risk attitude' to be a categorical or graded concept.

Risk-Attitudinal Unanimity Principle – qualitative version: *If all $v \in \mathcal{V}$ of non-zero correctness probability $Pr(v)$ are risk-neutral (-averse, -prone), then the meta-theory V is also risk-neutral (-averse, -prone).*

Risk-Attitudinal Unanimity Principle – quantitative version: *For all options $a \in A$, if all $v \in \mathcal{V}$ of non-zero correctness probability $Pr(v)$ assign the same risk premium $prem_v(p_a) = r$ to a's value prospect p_a, then the meta-theory assigns that same risk premium to a, i.e., $Prem_V(a) = r$.*

Why does the quantitative principle assume a unanimous risk premium for a's value prospect p_a rather than for a? The reason is simple: the $prem_v(a)$'s ($v \in \mathcal{V}$) are premia only for the empirical risk in a, while the $prem_v(p_a)$'s ($v \in \mathcal{V}$) are premia also for normative risk, since a's value prospect p_a also captures normative risk. By contrast, at the meta-level the principle uses the *option*-level risk premium $Prem_V(a)$, as $Prem_V(a)$ is already a premium for both types of risk, by being formed under normative uncertainty. Stating the principle using option-level risk premia at both levels – first-order and meta – would have been implausible: it would have required that a unanimous premium for *empirical* risk be adopted as the meta-theoretic premium for *empirical and normative* risk. The point becomes obvious if a is empirically riskless, i.e., yields a sure world: then a's first-order risk premia $prem_v(a)$ (v in $\mathcal{V}$) are unanimously zero, yet a non-zero meta-theoretic premium may be justified by normative uncertainty.

Of the four meta-theories, only the impartial theory achieves the principle:

[21] By the value prospect of an option a we here mean the theory-specific value prospect $p_{a,v}$ in case of a first-order theory v in $\mathcal{V}$, and the unconditional value prospect p_a in case of a meta-theory. Why does Definition 4 exclude options with riskless value prospect from its quantification? Requiring non-zero risk premia for essentially riskless options (i.e., requiring risk-aversion or -proneness towards such options) would be implausible, even for intuitively risk-averse or -prone theories.

Theorem 1 *The Risk-Attitudinal Unanimity Principle, in its qualitative and quantitative version, is satisfied by the impartial value theory IV, but can be violated by EV, FEV and DEV.*

Intuitively, the fully expectational theory violates the principle by being always risk-neutral; the expected and dual expected value theories violate it by respecting the first-order risk attitudes w.r.t. just empirical risk (EV) or just normative risk (DEV); and the impartial value theory satisfies it by subjecting all risk to the valuations in $\mathcal{V}$.

Often there is risk-attitudinal uncertainty, i.e., heterogeneity in the risk attitudes across first-order valuations. In this heterogeneous case, the impartial value theory forms a linear compromise between the competing first-order risk attitudes:

Theorem 2 *The degree of risk aversion of the impartial value theory IV towards an option $a \in A$ is the expected ('average') degree of risk aversion towards a's value prospect:*

$$Prem_{IV}(a) = \sum_{v \in \mathcal{V}} Pr(v) prem_v(p_a). \tag{1}$$

Intuitively, IV has an 'impartial' risk attitude in the sense of a linear compromise between first-order risk attitudes. IV thus satisfies Section 3's Risk-Attitudinal Impartiality Principle, *provided* that principles is given a suitable linear interpretation.

The meta-theoretic risk premium $Prem_{IV}(a)$ is a compromise between the prospect-level risk premia $prem_v(p_a)$ ($v \in V$), not the option-level risk premia $prem_v(a)$ ($v \in V$). This is a desirable feature, not a bug, because each $prem_v(a)$ only accounts for empirical risk in a, while each $prem_v(p_a)$ also accounts for normative risk.

6.2 Which meta-theories are based on the value prospect?

How do the new meta-theories perform w.r.t. the Value-Prospect Principle violated by the expected value theory? We give the answer in the next theorem, preceded by a formal re-statement of that principle.

Value-Prospect Principle: *Options $a, b \in A$ with same value prospect $p_a = p_b$ have same overall value $V(a) = V(b)$.*

Theorem 3 *The Value-Prospect Principle is satisfied by the impartial and fully expectational value theories IV and FEV, but can be violated by the expected and dual expected value theories EV and DEV.*

This positive result about IV holds trivially, as the impartial value of an option a is *by definition* determined by a's value prospect. The positive result about FEV holds because, as shown in the appendix, the fully expectational value of an option a equals the expectation of a's value prospect:

$$FEV(a) = \sum_{k \in \mathbb{R}} k p_a(k).$$

7 Where and when do the meta-theories agree?

The expected and dual expected value theories EV and DEV lie between the fully expectational theory FEV and the impartial theory IV:

- EV resembles FEV in its neutrality to normative risk, and IV in its impartial attitude to empirical risk.
- DEV resembles FEV in its neutrality to empirical risk, and IV in its impartial attitude to normative risk.

Loosely speaking, each of EV and DEV is somewhere risk-neutral (like FEV), and somewhere risk-attitudinally impartial (like IV). Two theorems will substantiate this claim, thereby substantiating the perfect duality or symmetry between EV and DEV. The set of empirically riskless options is

$$A_{\text{e-riskless}} = \{a \in A : a(x) = 1 \text{ for some world } x \in X\},$$

The set of options resulting in normatively riskless worlds is

$$A_{\text{n-riskless}} = \{a \in A : \text{ all } v \in \mathcal{V} \text{ s.t. } Pr(v) \neq 0 \text{ agree at all } x \in X \text{ s.t. } a(x) \neq 0\}.$$

Theorem 4 *The expected value theory EV matches*

(a) the fully expectational value theory FEV at options in $A_{e\text{-}riskless}$ (so is risk-neutral towards such options),

(b) the impartial value theory IV at options in $A_{n\text{-}riskless}$ (so has impartial risk attitude towards such options in the sense of the average risk premium (1)).

Theorem 5 *The dual expected value theory DEV matches*

(a) the fully expectational value theory FEV at options in $A_{n\text{-}riskless}$ (so is risk-neutral towards such options),

(b) the impartial value theory IV at options in $A_{e\text{-}riskless}$ (so has impartial risk attitude towards such options in the sense of the average risk premium (1)).

When do our meta-theories coincide at *all* options? This happens if risk-neutrality is certainly correct, i.e., only risk-neutral valuations have non-zero probability:

Theorem 6 *All four meta-theories EV, FEV, IV and DEV coincide globally if each valuation in $\mathcal{V}$ of non-zero probability is risk-neutral.*

This theorem even holds with an 'if and only if' in all but certain artificial cases.

8 Concluding remarks

The classic expected value theory EV is one of four 'expectational' meta-theories, marked by different attitudes to normative and empirical risk (Theorems 1, 4, 5), and distinct in whether the aggregation of value follows an ex-ante or ex-post approach w.r.t. normative or empirical risk. We have conditionally defended the impartial value theory

IV, on the grounds that it respects impartially the first-order risk attitudes (Theorem 2) and is based solely on value information (Theorem 3). Our defence is conditional in different ways. First, IV presupposes as usual that normative uncertainty can be quantified probabilistically and that value is numerically measurable and cross-theory comparable. Second, IV relies on a linear or expectational way to respect the first-order risk attitudes – it is 'linearly impartial'. While we are open to non-linear versions of IV, we insist that IV's linear form does not cause any risk-neutrality – as opposed to classic EV, whose linear form causes the questionable neutrality to normative uncertainty.

A Proofs

We here prove all results, in different order and based on many lemmas.

Lemma 1 *The fully expectational value of an option a in A is the expectation of its value prospect:*

$$FEV(a) = Exp(p_a) \ (= \sum_{k \in \mathbb{R}} k p_a(k)).$$

Proof. For all options a in A,

$$\begin{aligned} FEV(a) &= \sum_{(x,v) \in X \times \mathcal{V}} a(x) Pr(v) v(x) = \sum_{k \in \mathbb{R}} \sum_{(x,v) \in X \times \mathcal{V} : v(x) = k} a(x) Pr(v) k \\ &= \sum_{k \in \mathbb{R}} k \sum_{(x,v) \in X \times \mathcal{V} : v(x) = k} a(x) Pr(v) = \sum_{k \in \mathbb{R}} k p_a(k) = Exp(p_a). \ \blacksquare \end{aligned}$$

Proof of Theorem 3. IV and FEV satisfy the principle, by definition for IV and by Lemma 1 for FEV. EV and DEV can violate the principle: in Section 3.2's example, $p_b = p_c$ $(= 4_{50\%}0_{50\%})$ but $EV(b) = 2 \neq EV(c) = 1$ and $DEV(b) = 1 \neq DEV(c) = 2$. $\blacksquare$

Risk attitudes can be characterized in terms of evaluations of value prospects rather than options:

Lemma 2 *A valuation $v \in \mathcal{V}$ is risk-neutral (-averse, -prone) if and only if $v(p) = (<, >)\ Exp(p)$ for all value prospects p that are risky (i.e., do not assign probability one to any value).*

Proof. We just show the claim for risk-neutrality, since the claims for risk-aversion and -proneness are analogous. First, consider a risk-neutral $v \in \mathcal{V}$ and a risky value prospect p. We prove $v(p) = Exp(p)$. Pick an $a \in A$ such that $p_{a,v} = p$. By risk-neutrality, $v(a) = \sum_{x \in X} a(x) v(x)$. To show $v(p) = Exp(p)$, we prove $v(p) = v(a)$ and $Exp(p) = \sum_{x \in X} a(x) v(x)$. The former holds because $p = p_{a,v}$, and the latter because

$$\begin{aligned} Exp(p) &= \sum_{k \in \mathbb{R}} k p(k) = \sum_{k \in \mathbb{R}} k p_{a,v}(k) = \sum_{k \in \mathbb{R}} k \sum_{x \in X : v(x) = k} a(x) \\ &= \sum_{k \in \mathbb{R}} \sum_{x \in X : v(x) = k} a(x) k = \sum_{x \in X} a(x) v(x). \end{aligned}$$

Conversely, let $v(p) = Exp(p)$ for all risky value prospects p. Let $a \in A$ have risky value prospect $p_{a,v}$. We must show $v(a) = \sum_{x \in X} a(x)v(x)$. Letting $p = p_{a,v}$, this follows from the identity $v(p) = Exp(p)$, because, as in the first part of the proof, $v(p) = v(a)$ and $Exp(p) = \sum_{x \in X} a(x)v(x)$. ■

Lemma 3 *Every valuation $v \in \mathcal{V}$ is risk-neutral towards options $a \in A$ whose value prospect $p_{a,v}$ is riskless (i.e., assigns probability one to some value).*

Proof. Let $v \in \mathcal{V}$ and $a \in A$ such that $p_{a,v}(k) = 1$ for some $k \in \mathbb{R}$. We prove that v is risk-neutral towards a, i.e., that $v(a) = k$ $(= FEV(a))$. Pick any $x \in X$ such that $a(x) \neq 0$. Clearly, $v(x) = k$. Let b be the riskless option such that $b(x) = 1$. As $p_{a,v} = p_{b,v}$, we have $v(a) = v(b)$, by assumption on valuations in $\mathcal{V}$. Meanwhile $v(b) = v(x) = k$. So $v(a) = k$. ■

Proof of Theorem 2. Consider an option $a \in A$. Using Lemma 1,

$$FEV(a) = Exp(p_a) = Exp(p_a) \sum_{v \in \mathcal{V}} Pr(v) = \sum_{v \in \mathcal{V}} Pr(v) Exp(p_a).$$

Now

$$\begin{aligned} Prem_{IV}(a) &= FEV(a) - IV(a) \\ &= \sum_{v \in \mathcal{V}} Pr(v) Exp(p_a) - \sum_{v \in \mathcal{V}} Pr(v) v(p_a) \\ &= \sum_{v \in \mathcal{V}} Pr(v) \left[Exp(p_a) - v(p_a) \right] \\ &= \sum_{v \in \mathcal{V}} Pr(v) prem_v(p_a). \blacksquare \end{aligned}$$

Lemma 4 *If $a \in A_{n\text{-}riskless}$, then for all $v \in \mathcal{V}$ such that $Pr(v) \neq 0$ we have $p_{a,v} = p_a$ (whence $v(a) = v(p_a)$, applying v on both sides).*

Proof. Let $a \in A_{\text{n-riskless}}$. Then all $p_{a,v}$ with $v \in \mathcal{V}$ and $Pr(v) \neq 0$ coincide. Let p be that common value prospect. It equals p_a because, for all $k \in \mathbb{R}$,

$$\begin{aligned} p_a(k) &= \sum_{(x,v) \in X \times \mathcal{V} : v(x) = k} a(x) Pr(v) = \sum_{v \in \mathcal{V} : Pr(v) \neq 0} Pr(v) \sum_{x \in X : v(x) = k} a(x) \\ &= \sum_{v \in \mathcal{V} : Pr(v) \neq 0} Pr(v) \underbrace{p_{a,v}(k)}_{p(k)} = p(k) \sum_{v \in \mathcal{V} : Pr(v) \neq 0} Pr(v) = p(k) \times 1 = p(k). \blacksquare \end{aligned}$$

Proof of Theorem 4. (a) If $a \in A_{\text{e-riskless}}$, say $a(y) = 1$ where $y \in X$, then

$$FEV(a) = \sum_{(x,v) \in X \times \mathcal{V}} \underbrace{a(x)}_{=1(0) \text{ if } x=(\neq)y} Pr(v) \underbrace{v(x)}_{=v(a) \text{ if } x=y} = \sum_{v \in \mathcal{V}} Pr(v) v(a) = EV(a).$$

(b) Let $a \in A_{\text{n-riskless}}$. Then, assuming without loss of generality that $Pr(v) \neq 0$ for all $v \in \mathcal{V}$,

$$EV(a) = \sum_{v \in \mathcal{V}} Pr(v) v(a) = \sum_{v \in \mathcal{V}} Pr(v) v(p_a) = IV(a),$$

where the second identity uses Lemma 4. ■

Proof of Theorem 4. (a) Let $a \in A_{\text{n-riskless}}$. Then:

$$\begin{aligned} DEV(a) &= \sum_{(x,v)\in X\times\mathcal{V}} a(x)Pr(v)v(p_x) = \sum_{(x,v)\in X\times\mathcal{V}:a(x)\neq 0,Pr(v)\neq 0} a(x)Pr(v)v(p_x) \\ FEV(a) &= \sum_{(x,v)\in X\times\mathcal{V}} a(x)Pr(v)v(x) = \sum_{(x,v)\in X\times\mathcal{V}:a(x)\neq 0,,Pr(v)\neq 0} a(x)Pr(v)v(x). \end{aligned}$$

To prove $DEV(a) = FEV(a)$, it remains to show that $v(p_x) = v(x)$, assuming $a(x) \neq 0$ and $Pr(v) \neq 0$. As $a \in A_{\text{n-riskless}}$ and $a(x) \neq 0$, the world x (more precisely, the riskless option identified with x) belongs to $A_{\text{n-riskless}}$. So $v(p_x) = v(x)$ by Lemma 4.

(b) If $a \in A_{\text{e-riskless}}$, say $a(y) = 1$ where $y \in X$, then

$$DEV(a) = \sum_{(x,v)\in X\times\mathcal{V}} \underbrace{a(x)}_{=1(0) \text{ if } x=(\neq)y} Pr(v)v(p_x) = \sum_{v\in\mathcal{V}} Pr(v)v(p_y) = IV(a). \blacksquare$$

Proof of Theorem 1. Write RU_{qual} and RU_{quan} for the qualitative and quantitative versions of the Risk-Attitudinal Unanimity Principle. Let $\tilde{\mathcal{V}} = \{v \in \mathcal{V} : Pr(v) \neq 0\}$.

Claim 1: IV satisfies RU_{qual}.

Assume all $v \in \tilde{\mathcal{V}}$ are risk-averse; the proof is analogous for risk-neutrality or -proneness. Consider any $a \in A$ with risky value prospect p_a. We must show that IV is risk-averse towards a, i.e., that $IV(a) < FEV(a)$. Note

$$IV(a) = \sum_{v\in\mathcal{V}} Pr(v)v(p_a) = \sum_{v\in\tilde{\mathcal{V}}} Pr(v)v(p_a).$$

In the last expression, each $v(p_a)$ is below $Exp(p_a)$ by v's risk-aversion and Lemma 2. So

$$IV(a) < \sum_{v\in\tilde{\mathcal{V}}} Pr(v)Exp(p_a) = Exp(p_a)\sum_{v\in\tilde{\mathcal{V}}} Pr(v) = Exp(p_a) \times 1 = FEV(a),$$

where the last equality uses Lemma 1. This proves $IV(a) < FEV(a)$.

Claim 2: IV satisfies RU_{quan}.

This claim is a special case of Theorem 2, proved above.

Claim 3: FEV can violate RU_{qual} and RU_{quan}.

This claim is trivial. Just choose X, A, $\mathcal{V}$ and Pr such that the $v \in \tilde{\mathcal{V}}$ are all risk-averse (or all risk-prone); as FEV is risk-neutral, RU_{qual} is violated. If we moreover let all $v \in \tilde{\mathcal{V}}$ have same non-zero degree of risk aversion $prem_v(p_a)$ towards the value prospect p_a of some option $a \in A$ – e.g., by letting $\tilde{\mathcal{V}}$ be singleton – then RU_{quan} is also violated, because $Prem_{FEV}(a) = 0$.

Claim 4: EV can violate RU_{qual}.

Choose any X, A, $\mathcal{V}$ and Pr such that (i) all $v \in \tilde{\mathcal{V}}$ are risk-averse, and (ii) some world $y \in X$ is evaluated differently by at least two valuations in $\tilde{\mathcal{V}}$. We prove that EV is not risk-averse. Let a be the option which certainly yields y. So $a(x) = 1(0)$ if $x = (\neq)y$. Hence,

$$\sum_{x\in X} a(x)v(x) = v(y) = v(a) \text{ for all } v \in \mathcal{V}. \tag{2}$$

Now

$$\begin{aligned} EV(a) &= \sum_{v \in \mathcal{V}} Pr(v)v(a) \\ &= \sum_{v \in \mathcal{V}} Pr(v) \sum_{x \in X} a(x)v(x) \text{ by (2)} \\ &= \sum_{(x,v) \in X \times \mathcal{V}} a(x)Pr(v)v(x) = FEV(a). \end{aligned}$$

As $EV(a) = FEV(a)$, EV is risk-neutral towards a. So EV is not globally risk averse, noting that a's value prospect p_a is risky as a (i.e., the world y) is evaluated differently by different valuations in $\tilde{\mathcal{V}}$.

Claim 5: EV can violate RU_{quan}.

Choose any X, A, $\mathcal{V}$ and Pr such that there is a world $y \in X$ for which (i) $v(y)$ is not the same for all $v \in \tilde{\mathcal{V}}$, and (ii) all $v \in \tilde{\mathcal{V}}$ assign the same risk premium to y's value prospect, denoted $prem_v(p_y) \equiv prem(p_y)$. Such a choice is possible, namely by constructing a set of valuations $\mathcal{V}$ in three steps (and letting $Pr(v) \neq 0$ for all $v \in \mathcal{V}$): first, fix a $y \in X$ and fix how the $v \in \mathcal{V}$ evaluate worlds (riskless options), taking care that y is evaluated differently; second, fix a function $prem$ of value prospects p, where $prem(p)$ is zero if and only if p is riskless ($prem(p)$ will become the risk premium for p); third, extend each $v \in \mathcal{V}$ to risky options a by defining $v(a)$ as

$$v(a) = \sum_{x \in X} a(x)v(x) - prem(p_{a,v}) = Exp(p_{a,x}) - prem(p_{a,x}),$$

the difference between a's expected world value and a premium for the empirical risk. Each $v \in \mathcal{V}$ assigns to each value prospect p the value $v(p) = Exp(p) - prem(p)$ and hence the risk premium $prem_v(p) = Exp(p) - v(p) = prem(p)$ (which confirms the 'risk premium' interpretation given to the function $prem$).

Now let a be the riskless option which surely yields y. By (i), a's value prospect p_a is risky. Each $prem_v(p_a)$ is the same for all $v \in \mathcal{V}$, namely $prem(p_a)$. So RU_{quan} would require that $Prem_{EV}(a) = prem(p_a)$. Yet $Prem_{EV}(a) \neq prem(p_a)$, because $prem(p_a) \neq 0$ (as p_a is risky), while $Prem_{EV}(a) = 0$ (as EV is risk-neutral towards empirical-riskless options like a, by Theorem 4(a)).

Claim 6: DEV can violate RU_{qual}.

Choose any X, A, $\mathcal{V}$ and Pr such that some risk-averse $\tilde{v} \in \mathcal{V}$ is surely correct: $Pr(\tilde{v}) = 1$ (no normative uncertainty). So $\tilde{\mathcal{V}} = \{\tilde{v}\}$. Hence trivially all $v \in \tilde{\mathcal{V}}$ are risk-averse. So RU_{qual} requires of DEV to be risk-averse. But by Theorem 5(a) DEV globally coincides with the risk-neutral meta-theory FEV, as $A = A_{\text{n-riskless}}$.

Claim 7: DEV can violate RU_{quan}.

Choose X, A, $\mathcal{V}$ and Pr just as in Claim 6's proof. To see why RU_{quan} is violated, pick any $a \in A$ with risky value prospect p_a. As $\tilde{\mathcal{V}} = \{\tilde{v}\}$, trivially $prem_v(p_a)$ is the same for all $v \in \mathcal{V}$. So RU_{quan} requires of DEV that $Prem_{DEV}(a) = prem_{\tilde{v}}(p_a)$. Yet $prem_{\tilde{v}}(p_a) \neq 0$ (as $\tilde{v}$ is risk-averse and p_a is risky) while $Prem_{DEV}(a) = 0$ (as DEV is risk-neutral by the proof of Claim 6). ■

Proof Theorem 6. Let all $v \in \mathcal{V}$ with $Pr(v) \neq 0$ be risk-neutral. Consider any option $a \in A$. We must show that $EV(a) = FEV(a) = IV(a)$. As before, write $\tilde{\mathcal{V}} = \{v \in \mathcal{V} : Pr(v) \neq 0\}$. Now each $v \in \tilde{\mathcal{V}}$ is risk-neutral towards a, by risk-neutrality of v (complemented by Lemma 3 in case a's value prospect $p_{a,v}$ is riskless). So,

$$v(a) = \sum_{x \in X} a(x)v(x) \text{ for all } v \in \tilde{\mathcal{V}}. \tag{3}$$

To see why $EV(a) = FEV(a)$, note first that

$$EV(a) = \sum_{v \in \mathcal{V}} Pr(v)v(a) = \sum_{v \in \tilde{\mathcal{V}}} Pr(v)v(a) = \sum_{v \in \tilde{\mathcal{V}}} Pr(v) \sum_{x \in X} a(x)v(x),$$

where the last identity uses (3). Rearranging,

$$EV(a) = \sum_{(x,v) \in X \times \tilde{\mathcal{V}}} a(x)Pr(v)v(x) = \sum_{(x,v) \in X \times \mathcal{V}} a(x)Pr(v)v(x) = FEV(a).$$

To see why $IV(a) = EV(a)$, note first that

$$IV(a) = \sum_{v \in \mathcal{V}} Pr(v)v(p_a) = \sum_{v \in \tilde{\mathcal{V}}} Pr(v)v(p_a).$$

In the last expression each $v(p_a)$ reduces to $Exp(p_a)$ by risk-neutrality of v via Lemma 2. So,

$$IV(a) = \sum_{v \in \tilde{\mathcal{V}}} Pr(v)Exp(p_a) = Exp(p_a) \sum_{v \in \tilde{\mathcal{V}}} Pr(v) = Exp(p_a) \times 1 = FEV(a),$$

where the last identity uses Lemma 1. ■

References

Barry, C., Tomlin, P. (2016) Moral uncertainty and permissibility: evaluating option sets, *Canadian Journal of Philosophy* 46: 898-923

Bossert, W., Weymark, J. (2004) Utility in social choice. In: Barberà, S., Hammond P., Seidl, C. (eds.) *Handbook of utility theory, vol. 2: Extensions*, Kluwer, Dordrecht, pp. 1099-1177

Bradley, R., Stefánsson, O. (2017) What Is Risk Aversion? *The British Journal for the Philosophy of Science*

Bradley, R., Drechsler, M. (2014) Types of Uncertainty, *Erkenntnis* 79(6):1225-1248

Broome, J. (1991) *Weighing Goods*, Oxford: Blackwell

Buchak, L. (2013) *Risk Aversion and Rationality*, Oxford University Press

Diamond, P. A. (1967) Cardinal welfare, individual ethics, and interpersonal comparison of utility: comment, *Journal of Political Economy* 75: 765-6

Dietrich, F., List, C. (2013) A reason-based theory of rational choice, *Noûs* 47: 104-134

Dietrich, F., List, C. (2017) What matters and how it matters: a choice-theoretic representation of moral theories, *Philosophical Review* 126: 421-479

Fleurbaey, M. (2010) Assessing risky social situations, *Journal of Political Economy* 118: 649-680

Fleurbaey, M., Voorhoeve, A. (2016) Priority or equality for possible people? *Ethics* 126: 929-954

Fleurbaey, M., Zuber, S. (2017) Fair management of social risk, *Journal of Economic Theory* 169: 666-706

Greaves, H., Ord, T. (2017) Moral uncertainty about population ethics, *Journal of Ethics and Social Philosophy* 12: 135-167

Harsanyi, J. (1978) Bayesian decision theory and utilitarian ethics, *American Economic Review* 68: 223-228.

Jackson, F., Smith, M. (2006) Absolutist Moral Theories and Uncertainty, *Journal of Philosophy* 103: 267-283

Lazar, S. (2017) Deontological Decision Theory and Agent-Centred Options, *Ethics* 127: 579-609

Lockhart, T. (2000) *Moral Uncertainty and its Consequences*, Oxford University Press

MacAskill, W. (2014) *Normative Uncertainty*, Doctoral thesis, University of Oxford

MacAskill, W., Ord, T. (forth.) Why maximize expected choice-worthiness, *Noûs*

McCarthy, D. (2006) Utilitarianism and prioritarianism I, *Economics and Philosophy* 22:335-363

McCarthy, D. (2008) Utilitarianism and prioritarianism II, *Economics and Philosophy* 24: 1-33

Nissan-Rozen, I. (2015) Against Moral Hedging. *Economics and Philosophy* 31: 1-21

Oddie, G. (1994) Moral uncertainty and human embryo experimentation. In: Fulford, K. W. M., Gillett, G., Soskice, J. M. (eds.) *Medicine and Moral Reasoning*, Cambridge University Press, pp. 3-144

Ross, J. (2006) Rejecting Ethical Deflationism, *Ethics* 116: 742-68

Savage, L. J. (1954) *The Foundations of Statistics*, New York: Wiley

Sepielli, A. (2006) Ted Lockhart, Moral Uncertainty and Its Consequences: moral uncertainty and its consequences. *Ethics* 116: 601-604

Sepielli, A. (2009) What to Do When You Don't Know What To Do. In: Shafer-Landau, R. (ed.) *Oxford Studies in Metaethics*, Oxford University Press, p. 35

Sepielli, A. (2017) How moral uncertaintism can be both true and interesting, *Oxford Studies in Normative Ethics* 7

Tenenbaum, S. (2017) Action, Deontology, and Risk: Against the Multiplicative Model, *Ethics* 127: 674-707

von Neumann, J., Morgenstern, O. (1944) *Theory of Games and Economic Behavior*, Princeton University Press

Weatherson, B. (2014) Running Risks Morally, *Philosophical Studies* 167(1): 141-63

Weirich, P. (1986) Expected utility and risk, *The British Journal for the Philosophy of Science* 37(4): 419-442

Weirich, P. (2004) *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances*, Oxford University Press

Weymark, J. (1991) A reconsideration of the Harsanyi–Sen debate on utilitarianism. In Elster, J., Roemer, J. E. (eds.) *Interpersonal Comparisons of Well-Being*, Cambridge University Press, pp. 255

Williams, J. R. G. (2017) Indeterminate Oughts, *Ethics* 127: 645-673