



HAL
open science

Expected Value Under Normative Uncertainty

Franz Dietrich, Brian Jabarian

► **To cite this version:**

Franz Dietrich, Brian Jabarian. Expected Value Under Normative Uncertainty. 2020. halshs-02905431v2

HAL Id: halshs-02905431

<https://shs.hal.science/halshs-02905431v2>

Submitted on 17 Mar 2021 (v2), last revised 13 Oct 2021 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CES

Centre d'Économie de la Sorbonne
UMR 8174

Expected Value Under Normative Uncertainty

Franz DIETRICH, Brian JABARIAN

2020.15R

Version révisée



Expected Value Under Normative Uncertainty

This version: March 2021¹

Franz Dietrich
Paris School of Economics & CNRS

Brian Jabarian
University of Paris 1

Abstract

Maximising expected value is the classic doctrine in choice theory under empirical uncertainty, and a prominent proposal in the emerging philosophical literature on normative uncertainty, i.e., uncertainty about values. But how should Expectationalism be stated in general, when we can face both uncertainties simultaneously, as is common in life? Surprisingly, different possibilities arise, ranging from Ex-Ante to Ex-Post Expectationalism, with several hybrid versions. Expectationalism thus faces the classic dilemma between ex-ante and ex-post approaches. Different expectational theories reach diverging evaluations, use different modes of reasoning, and take different attitudes to risk; but they converge under an interesting (necessary and sufficient) condition. We relate choice under normative uncertainty to choice by Harsanyi’s ‘impartial observer’ and Rawls’s person behind the ‘veil of ignorance’, who are uncertain about their own identity and values; and we relate Expectationalism under normative uncertainty to Harsanyi’s linear approach to aggregating values of different individuals or identities.

1 The problem

When evaluating choice options, we often face two types of uncertainty. Decision theory has focused on *empirical* uncertainty: uncertainty about empirical facts, such as facts about weather or election outcomes. Philosophers have recently turned to *normative* uncertainty: uncertainty about the evaluation itself, for instance because of competing normative intuitions. Important contributions include Oddie (1994), Lockhart (2000), Jackson and Smith (2006), MacAskill (2014, 2016), Bradley and Drechsler (2014), Sepielli (2009), Weatherson (2014), Lazar (2017), Greaves and Cotton-Barratt (2019), Tarsney (2018a, 2019), Podgorski (2019), MacAskill and Ord (2020), Riedener (2020), and Dietrich and Jabarian (2021a). Normative uncertainty is omnipresent in deliberation and decision-making: parents may wonder how much they should value

¹Personal acknowledgments to be added. Franz Dietrich acknowledges support by the French National Research Agency through three grants (ANR-17-CE26-0003, ANR-16-FRAL-0010 and ANR-17-EURE-0001). Brian Jabarian acknowledges support by the Calouste Gulbenkian Foundation under a Global Excellence Scholarship.

child autonomy, even when certain about all relevant empirical facts; agents may wonder which consequences matter most, and whether deontological considerations matter too; reason-based agents (as in Dietrich and List 2013, 2017) may wonder which properties matter, and how they matter; and so on.

Much can be debated about the metaphysical status of normative uncertainty: Is it uncertainty about subjective or objective facts? About real or constructed facts? And so forth. We put these important debates aside. But we stress two things: normative uncertainty is meaningful under many interpretations and metaethical views, and it is not formally reducible to standard choice-theoretic uncertainty.²

A prominent proposal in the theory of normative uncertainty is to maximise expected value across standards of evaluation, an approach pioneered in some of the cited works, by Oddie, MacAskill, and Ord, respectively. Riedener (2020) provides a sophisticated axiomatic defence. Much of today’s debate focuses on whether this approach – to be called *Expectationalism* – is justified. Expectationalism has for instance been defended through analogies with empirical uncertainty and through formal arguments (e.g., Riedener’s work), but criticized for relying on a precise quantification of normative uncertainty and on certain measurements and comparisons of value (e.g., Tarsney 2018b). Instead of addressing the important question ‘Expectationalism or not?’, we shall assume Expectationalism, and ask a new question: ‘Expected value of what?’ A particular answer was so far taken for granted: expected value *of the option*. We call this Standard Expectationalism. Standard Expectationalism is hybrid in nature: it reasons ‘empirically ex-ante’, but ‘normatively ex-post’, roughly speaking because the objects of expected value (the ‘options’) contain only empirical uncertainty. There are two non-hybrid incarnations of Expectationalism, Ex-Ante Expectationalism and Ex-Post Expectationalism, which form the expected value of prospects that respectively contain *all* uncertainty (empirical or normative) or *no* uncertainty. Between these two extremes, there is a spectrum of more or less ex-post versions of Expectationalism.

The dilemma between ex-ante and ex-post reasoning is prominent in other fields of formal ethics and aggregation theory; see Diamond (1976), McCarthy (2006, 2008, 2015), Fleurbaey (2010), Fleurbaey and Voorhoeve (2016), and Fleurbaey and Zuber (2017). The theory of normative uncertainty cannot escape this dilemma. Just as social egalitarians face a dilemma between ex-ante and ex-post equality, so expectationalists in individual decision theory face a dilemma between ex-ante, ex-post, and hybrid formulations of Expectationalism. The dilemma always takes the same form: should the given paradigm or objective – for instance equality or (in our case) expected value – be pursued from an ex-ante or ex-post perspective? The ex-ante/ex-post dilemma reaches also into non-expectational approaches to normative uncertainty, such as max-min approaches, of which one could again envision ex-ante, ex-post, and hybrid versions.

²Some decision-theoretic models can be reinterpreted in terms of normative uncertainty. Examples are multi-utility models and Harsanyi’s (1978) impartial-observer model, to which we come in Section 7. But the attempt to simply re-interpret choice-theoretic risk or uncertainty (in von-Neumann-Morgenstern’s 1944 or Savage’s 1954 framework) normatively runs into formal problems. For instance, writing ‘normative information’ into Savage’s states implies letting states determine utilities, in ways not even compatible with standard state-dependent utility theory (see however Riedener 2020 for a non-standard analysis). This would ultimately undermine the two-attitude make-up of choice theory, which is based on two independent ingredients, beliefs and values (or formally, probabilities and utilities).

But we here exclusively focus on Expectationalism.

In putting the spectrum of expectational theories on the table, we raise the problem of the very meaning and formulation of Expectationalism. We will not defend some particular version of Expectationalism, i.e., some degree of ex-post reasoning. Deciding between the rival expectational theories or reasoning modes is an important future problem, which might prove as difficult as deciding between ex-ante and ex-post egalitarianism.

The paper has a simple structure. Section 2 introduces the simplest possible formal framework for capturing normative and (where needed) empirical uncertainty. Section 3 introduces four salient expectational theories – four solutions to the problem ‘expectation of what?’. Section 4 defines these four theories formally. Section 5 illustrates how they reach diverging evaluations and take different attitudes to risk. Section 6 defines Expectationalism in general, going beyond the four special theories. Section 7 relates Expectationalism to linear welfare aggregation in a group, by a social planner or a Harsanyi-type ‘impartial observer’. Section 8 concludes the paper.

2 A framework of normative and empirical uncertainty

We now present and discuss the formal ingredients one by one.

The objects of evaluation. The agent – ‘you’ – considers a non-empty set A of objects, called ‘options’. They could be policy measures, social arrangements, income distributions etc. So far we leave open whether options contain empirical risk.

Competing valuations. You are uncertain about the correct value of options. Value is taken to be an absolute (cardinal) rather than comparative (ordinal) notion. So, a possible standard of evaluation is given by a value function rather than order, formally a real-valued function v on A , called a *valuation*, assigning to each option a in A a value $v(a)$. Let \mathcal{V} be the set of valuations you deem possible, formally a finite non-empty set of functions from A to \mathbb{R} .

Let us give examples, which will involve either moral valuations (for moral choice) or rational valuations (for rational choice). \mathcal{V} could contain a utilitarian, an egalitarian, and some deontological valuation. Or in intertemporal choice, \mathcal{V} could contain a valuation with exponential discounting and another with hyperbolic discounting. Or \mathcal{V} could consist of valuations of same type differing only in a parameter, such as: prioritarian social-welfare valuations of different degrees of prioritarianism, or egalitarian social-welfare valuations of different degrees of inequality-aversion, or valuations of intertemporal well-being of different discounting of future well-being, or (for risky options) valuations of different degrees of risk-aversion, etc. In such parametric examples, normative uncertainty boils down to uncertainty about the correct parameter value: the correct amount of prioritarianism, inequality-aversion, discounting, risk-aversion, etc.³

³Our examples show that normative uncertainty comes in two species: mere ‘parameter uncertainty’ and fundamental ‘model uncertainty’. A similar distinction is made in other fields, especially statistics and macroeconomics (e.g., Hansen and Sargent 2001).

Credences in valuations. You assign to each valuation v in \mathcal{V} a correctness probability $Pr(v) \geq 0$, where $\sum_{v \in \mathcal{V}} Pr(v) = 1$. Probabilities capture degrees of belief about value.

Meta-value. Given these credences, how should you evaluate options *overall*? An answer takes the form of a *meta-valuation* or *theory*, formally a function assigning to each option in A a (meta-)value in \mathbb{R} . To distinguish meta-valuations from valuations, we denote them by upper-case letters like ‘ V ’. Two examples suffice for now. Standard Expectationalism evaluates options $a \in A$ by their expected value: $V(a) = \sum_{v \in \mathcal{V}} Pr(v)v(a)$. Another (decidedly non-expectational) theory evaluates options $a \in A$ by their minimal possible value: $V(a) = \min_{v \in \mathcal{V}: Pr(v) \neq 0} v(a)$.

Some would call our theories ‘meta-theories’, and our valuations ‘first-order theories’. Moreover, we would have defined theories as value *functions* rather than value *orders* on A . Readers who prefer an ordinal notion of meta-value can replace each meta-valuation by the order it induces on A .

Measurability and comparability of value. As usual in the expectational approach, we take first-order value to be measurable and comparable across valuations. Full measurability makes it meaningful to say that an option x has value 7 under a valuation v ($v(x) = 7$), or is twice as valuable as another option y ($v(x) = 2v(y)$), or exceeds z ’s value by 2 ($v(x) - v(z) = 2$), etc. Full comparability makes it meaningful to say that two valuations v and v' assign same value to option x ($v(x) = v'(x)$), or same value gain to the change from option x to option y ($v(y) - v(x) = v'(y) - v'(x)$), etc.⁴ Such assumptions are strong and debatable. They can be relaxed, in ways that differ across versions of Expectationalism.⁵ We set aside when and how measurability and comparability can be justified,⁶ and how they could be relaxed by different versions of Expectationalism.

Adding empirical uncertainty: options as lotteries. The above framework is complete as a model of purely normative uncertainty; empirical uncertainty in options is allowed, but not modelled. To add empirical uncertainty explicitly, we hereafter assume that options in A are lotteries on a given set X of *outcomes*, i.e., functions a from X to $[0, 1]$ such that $\sum_{x \in X} a(x) = 1$, where $a(x)$ is non-zero for only finitely many x in X . An option is *riskless* if some outcome has probability one, and *risky* otherwise. Outcomes represent empirical states of affairs after resolution of empirical uncertainty. Under the simplest interpretation, outcomes are ‘consequences’ of actions; this limits us to (normative uncertainty between) consequentialist valuations. But outcomes could be broader than ‘consequences’. They could for instance encompass intentions and/or the choice context. This allows for non-consequentialist valuations.⁷

⁴Comparability and measurability are addressed by Bossert and Weymark (2004), and in the context of normative uncertainty by, e.g., Ross (2006), Sepielli (2009) and Tarsney (2018b).

⁵For instance, all versions need only *affine* measurements of value, and Standard Expectationalism needs only *unit* comparisons, not *level* comparisons.

⁶Justifying cross-valuation comparisons is easier if \mathcal{V} consists of theories of similar type, e.g., egalitarian theories with different degrees of inequality-aversion.

⁷Outcomes should contain everything that could bear value according to the agent we wish to model; only then can we faithfully model the agent’s normative uncertainty. If the agent believes the

Not all lotteries on X need to count as options, i.e., belong to A . But let A contain at least the riskless lotteries, which assign probability one to some outcome. We occasionally apply valuations v to outcomes x rather than options; of course, $v(x)$ stands for the value $v(a)$ of the riskless option a corresponding to x .

Economists distinguish between ‘risk’, in which probabilities are (in some sense) objective, and ‘uncertainty’, in which they are subjective. In our model, all probabilities – of outcomes or valuations – are exogenously given. Technically, this makes our framework one of (empirical and normative) risk, not (empirical and normative) uncertainty. But one can *interpret* probabilities subjectively. Regarding valuations, our term ‘credence’ already suggest a subjective interpretation. Regarding outcomes, one can interpret $a(x)$ as the agent’s subjective probability of outcome x under option a .⁸

Valuations of vNM type and of non-vNM type. A valuation v in \mathcal{V} could have the notorious von-Neumann-Morgenstern property, i.e., be of ‘vNM type’, as we shall say. A valuation v is of *vNM-type* if it evaluates each option by the expected value of its outcome, formally

$$v(a) = \sum_{x \in X} a(x)v(x) \text{ for all options } a \in A.$$

Valuations of vNM-type could also be called ‘expectational’, but in a different sense from that of expectational *meta*-valuations.⁹ Ever since the Harsanyi-Sen debate, it is controversial whether the vNM property is a necessary property of a coherent valuation, or (as we believe) a coincidental property that may or not hold.¹⁰

Our model is ecumenical: you (the agent) could be utterly certain that correct value is of vNM-type, by having positive credence only in vNM-type valuations; or be utterly certain of the opposite, by having positive credence only in non-vNM-type valuations; or be uncertain about the issue, by having positive credence in both types of valuation. Being ecumenical is important, because even if we (as modellers) were certain that true

context could matter normatively, then outcomes cannot exclude the context. Taking outcomes to be consequences limits us to consequentialist agents: agents who are certain that the correct valuation is consequentialist. If instead outcomes go beyond consequences, then a valuation v may or not be consequentialist. It is consequentialist if the value of options is fully determined by their consequence aspects, i.e., if $v(a) = v(b)$ for all options $a, b \in A$ that contain the same consequences (but possibly different contexts or other non-consequence features). Normative uncertainty between non-consequentialist valuations is addressed by Barry and Tomlin (2016) and Tenenbaum (2017).

⁸Subjective probabilities may be unobservable, which is why orthodox economists in an empiricist tradition feel uncomfortable with having subjective probabilities as model primitives, rather than as ‘outputs’ of characterization theorems such as Savage’s Theorem. Being not committed to empiricism, we do not mind subjective probabilities as model primitives.

⁹A vNM-type valuation is ‘expectational’ in its response to *empirical* uncertainty – the only sort of uncertainty faced by a first-order valuation. In this paper, ‘Expectationalism’ refers to a meta-valuation approach in the face of *normative* uncertainty.

¹⁰The Harsanyi-Sen debate is a debate about whether an ethically relevant notion of utility, especially one relevant to utilitarianism, is of vNM type, as is defended by Harsanyi and rejected by Sen. Broome (1991), Weymark (1991), Nissan-Rozen (2015), Fleurbaey and Mongin (2016), and Greaves (2017) present diverging analyses of the debate. Harsanyi’s (1955) famous theorem on additive social welfare is based on vNM-type utility. Depending on whether utility in the utilitarian sense is vNM, this theorem does or not pertain to utilitarianism.

value is of some type, the agent we model might not share this certainty. We should for instance not assume that \mathcal{V} contains only vNM-type valuations or only non-vNM-type valuations; this would restrict the model to very special agents who are *certain* that value is of vNM-type or *certain* that value is not of this type, unlike most or all philosophers interested in the vNM property.

A deeper methodological issue is at stake. The field of normative uncertainty is engaged in meta-normativity. It should thus avoid prejudging first-order normative questions. It should take people's *actual* normative beliefs and uncertainties at face value, however non-ideal or 'mistaken' they might be, and tell people how to ideally respond to their non-ideal normative beliefs. Even if true value were necessarily of vNM type, as some philosophers argue, then we should not assume the agent is aware of this. This said, the (very common) restriction to vNM-type valuations can sometimes be legitimate as a working assumption. We avoid this restriction.

3 Expected value of what?

Expectational theories evaluate options by the expected value of *some object*. That object is the prospect offered by the option, but there are different types of prospect: the ex-ante prospect, the ex-post prospect, and hybrid prospects from some intermediate (ex-interim) perspective.

Think of prospects as probability distributions. More precisely, one can define prospects equivalently as distributions over empirical-normative worlds ('world prospects') or distributions over resulting value levels ('value prospects'). We shall later only work with value prospects. But let us start with world prospects. An *empirical-normative world* – for short, a *world* – is a pair (x, v) of an outcome in X (an 'empirical world') and a valuation in \mathcal{V} (a 'normative world'). In a world, all empirical or normative uncertainty is resolved. A *world prospect* is a probability distribution over worlds, representing how likely worlds are (where for simplicity only finitely many worlds have non-zero probability). Each option a generates an (ex-ante) world prospect, under which the probability of a world (x, v) is the product $a(x)Pr(v)$ of the probabilities of outcome x (under option a) and valuation v . This world prospect is ex-ante because no uncertainty is resolved; ex-post and hybrid world prospects will be defined shortly.

We can now give four possible answers to the question 'Expected value of what?', hence four ways to reason towards the meta-value of a given option a . We keep the four answers informal; formal definitions follow in Section 4.

- *Normatively ex-post reasoning*: You place yourself in a normatively ex-post and empirically ex-ante position, by considering a given valuation v and the lottery of empirical outcomes generated by option a . So you face the *normatively ex-post* world prospect, in which v has (marginal) probability one and any outcome x in X has (marginal) probability $a(x)$. It yields the value $v(a)$. Stepping outside this position, you then form the expectation of the value $v(a)$ across valuations v in \mathcal{V} . This is Standard Expectationalism.
- *Ex-post reasoning*: You place yourself in a fully ex-post position, by considering a given outcome x and a given valuation v . So you face the *ex-post* world prospect,

in which world (x, v) has probability one. It yields the value $v(x)$. Stepping outside this position, you then form the expectation of the value $v(x)$ across worlds (x, v) in $X \times \mathcal{V}$. This is Ex-Post Expectationalism.

- *Ex-ante reasoning*: You place yourself in the fully ex-ante position, in which both parts of the empirical-normative world are unknown. So you face the *ex-ante* world prospect, defined above. You then form the expected value of this ex-ante prospect; how this works is shown in Section 4. This is Ex-Ante Expectationalism.
- *Empirically ex-post reasoning*: You place yourself in an empirically ex-post and normatively ex-ante position, by considering a given outcome x and the probability distribution over valuations Pr reflecting your normative uncertainty. So you face the *empirically ex-post* world prospect, in which x has (marginal) probability of one and any valuation v has (marginal) probability $Pr(v)$. You then form the expected value of this world prospect, in a way shown in Section 4. This is Reverse Expectationalism. It is the reverse or ‘dual’ of Standard Expectationalism, as it reasons ex-ante where Standard Expectationalism reasons ex-post, and vice versa.

	normatively ex-post	normatively ex-ante
empirically ex-post	Ex-Post Expectationalism	Reverse Expectationalism
empirically ex-ante	Standard Expectationalism	Ex-Ante Expectationalism

Table 1: Four expectational theories and their modes of reasoning

These four answers to the question ‘Expected value of what?’ were given in the form of world prospects, but they can be redescribed as *value* prospects. Value prospects are prospects of achieving certain value levels (not worlds) with certain probabilities, for instance achieving value 4 with probability 1/2 and value 0 with probability 1/2. A world prospect immediately induces a value prospect (mathematically, by taking the image of the world prospect under the mapping $(x, v) \mapsto v(x)$ from worlds to resulting values). For instance, the ex-post world prospect under which world (x, v) is certain induces the riskless value prospect under which the value $v(x)$ is certain.

Formally, a *value prospect* is simply a lottery over real numbers, 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 in \mathbb{R} have non-zero probability $p(k)$. Each option a generates a value prospect, denoted p_a . It reflects empirical and normative uncertainty, as the resulting value $v(x)$ depends on both x and v , hence on the empirical-normative world (x, v) . The probability that the resulting value is (say) 4 is the sum-total probability of all worlds (x, v) such that $v(x) = 4$. The just-defined value prospect p_a of an option a is an ex-ante construct: no uncertainty is yet resolved. Indeed, p_a is simply the value prospect induced by the ex-ante world prospect. Partly or fully ex-post value prospects are definable by eliminating one or both sources of uncertainty.

We now define the four kinds of value prospect formally. They correspond exactly to the four kinds of world prospect above, respectively:¹¹

- The **(ex-ante) value prospect of option** $a \in A$ is the value prospect ‘ p_a ’ such

¹¹Compare our value prospects with Rowe and Voorhoeve’s (2018) well-being prospects in a context of health ethics under (purely empirical) risk, uncertainty, or ambiguity.

that any value $k \in \mathbb{R}$ has probability

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

- The **(normatively ex-post) value prospect of option** $a \in A$ **given valuation** $v \in \mathcal{V}$ is the value prospect ' $p_{a,v}$ ' such that any value $k \in \mathbb{R}$ has probability

$$p_{a,v}(k) = \text{'probability that } a \text{ leads to value } k \text{ given } v\text{'} = \sum_{x \in X: v(x)=k} a(x).$$

- The **(empirically ex-post) value prospect given outcome** $x \in X$ is the value prospect ' p_x ' such that any value $k \in \mathbb{R}$ has probability:

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

- The **(ex-post) value prospect given** $x \in X$ **and** $v \in \mathcal{V}$ is the riskless value prospect ' $p_{x,v}$ ' under which the value is $v(x)$ with probability one.¹²

4 Four expectational theories

We now formally define the four expectational theories discussed in Section 3. Each takes the expected value of a certain prospect, as will be clear either by definition or by Theorem 1 below. We begin with the two theories whose definitions do not explicitly refer to prospects.

Standard Expectationalism (' EV_{stan} '): *The meta-value of an option $a \in A$ is the expected value of the option itself:*

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

This theory reasons empirically ex-ante, because the object whose expected value it forms (the option) contains empirical risk. The second theory reasons ex-post: it forms the expected value of *the outcome*, which no longer contains empirical risk. This requires averaging across both outcomes and valuations, hence across empirical-normative worlds (x, v) . Formally:

Ex-Post Expectationalism (' EV_{post} '): *The meta-value of an option $a \in A$ is the expected value of the outcome:*

$$EV_{post}(a) = \sum_{(x,v) \in X \times \mathcal{V}} \underbrace{a(x)Pr(v)}_{\text{prob. of } (x,v)} v(x).$$

The third theory reasons fully ex-ante. It operates neither at the fully ex-post level of outcomes, nor at the empirically ex-post level of options, but at the level of ex-ante

¹²The value prospects p_x and $p_{x,v}$ can be regarded as special cases of the value prospects p_a and $p_{a,v}$, respectively. Just choose a to be the riskless option that yields x for sure.

value prospects. But how can a valuation v in \mathcal{V} evaluate value prospects rather than options, i.e., how should we define $v(p)$ for a value prospect p ? We pick any option a in A whose value prospect given v is p , and identify $v(p)$ with $v(a)$. If for instance p is the value prospect ‘the value is 1 or 0 equiprobably’, then we pick an option a which equiprobably has some outcome x of value $v(x) = 1$ or some outcome y of value $v(y) = 0$, and define $v(p)$ as $v(a)$. Formally, the *value of a value prospect p under a valuation v* in \mathcal{V} – denoted $v(p)$ – is the value $v(a)$ of any option $a \in A$ such that $p_{a,v} = p$. This definition implicitly rests on an assumption that we make for the rest of the paper:

Assumption: For each valuation v in \mathcal{V} and value prospect p , (i) A contains an option a whose value prospect given v , $p_{a,v}$, is p ; and (ii) any two such options a in A have same value $v(a)$.

Condition (i) is a typical richness assumption: the set of options A should be sufficiently inclusive, i.e., contain options with any given value prospects. Condition (ii) is a consistency assumption on the valuations in \mathcal{V} . It is compatible with most or all first-order theories one would naturally want to consider.

Ex-Ante Expectationalism (EV_{ante}): *The meta-value of an option $a \in A$ is the expected value of the ex-ante prospect:*

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

Note an intended discrepancy: in $v(p_a)$, a *given* valuation (v) is applied to a prospect (p_a) that is formed *ex-ante*, when the valuation was uncertain rather than given. Precisely this is what ex-ante reasoning should do, as it should ask how attractive each *ex-ante* prospect is on average across possible valuations.

The fourth theory calculates the expected value of yet another object: the empirically ex-post value prospect. This requires averaging across outcomes and valuations, hence across empirical-normative worlds (x, v) .

Reverse Expectationalism (EV_{rev}): *The meta-value of an option $a \in A$ is the expected value of the empirically ex-post prospect:*

$$EV_{rev}(a) = \sum_{(x,v) \in X \times \mathcal{V}} \underbrace{a(x)Pr(v)}_{prob. \text{ of } (x,v)} v(p_x).$$

This theory reverses the reasoning of Standard Expectationalism: it reasons empirically ex-post rather than normatively ex-post.

The following theorem re-expresses the four theories in a unified format, showing that they only differ in the ‘locus’ of expectation-taking, i.e., in the sort of prospect whose expected value they take.

Theorem 1 Each expectational theory $V \in \{EV_{ante}, EV_{post}, EV_{stan}, EV_{rev}\}$ evaluates any option $a \in A$ by the expected value of a specific value prospect, i.e.,

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

where ‘ p ’ is a place holder for the

- ex-ante value prospect p_a if $V = EV_{ante}$,
- ex-post value prospect $p_{x,v}$ if $V = EV_{post}$,
- normatively ex-post value prospect $p_{a,v}$ if $V = EV_{stan}$,
- empirically ex-post value prospect p_x if $V = EV_{rev}$.

5 Illustration of these four expectational theories and their risk attitudes

Suppose you hesitate between just two valuations, v and v' . You have credence $\frac{1}{2}$ in each of them, and credence 0 in all other valuations (if any) in \mathcal{V} . Both valuations v and v' are risk-averse: they penalise options for any uncertainty about how good their outcome is (what this means will become clear). So you are sure that risk-aversion is correct, i.e., that the correct valuation is risk-averse. You now compare two options. Both options lead to the value prospect ‘value 4 with probability $\frac{1}{2}$, value 0 with probability $\frac{1}{2}$ ’, denoted $4_{50\%}0_{50\%}$, but for very different reasons:

- Option 1 involves only normative risk. It surely has some outcome x , whose value is either $v(x) = 4$ or $v'(x) = 0$.
- Option 2 involves only empirical risk. It has either some outcome y or some outcome z (equiprobably), where it is uncontroversial between v and v' that y has value 4 and z has value 0. As v and v' are risk-averse, they evaluate the option below the expected resulting value of $\frac{1}{2}4 + \frac{1}{2}0 = 2$. Let them assign the value be 1 to the option. The gap from 1 to 2 is the ‘risk penalty’ or ‘risk premium’.

	value prospect			evaluation by					
	given v	given v'	ex-ante	v	v'	EV_{stan}	EV_{post}	EV_{ante}	EV_{rev}
option 1	$4_{100\%}$	$0_{100\%}$	$4_{50\%}0_{50\%}$	4	0	2	2	1	1
option 2	$4_{50\%}0_{50\%}$	$4_{50\%}0_{50\%}$	$4_{50\%}0_{50\%}$	1	1	1	2	1	2

Table 2: Four expectational theories applied to two concrete options

Table 2 displays the (ex-ante and normatively ex-post) value prospects of the options and the evaluations by both first-order theories and the four expectational meta-theories. The four meta-evaluations are obtained as follows:

- *Standard* Expectationalism forms the average value of the option. This yields $\frac{1}{2}4 + \frac{1}{2}0 = 2$ or $\frac{1}{2}1 + \frac{1}{2}1 = 1$, respectively.
- *Ex-Post* Expectationalism forms the average value of the outcome. Recall that this in principle requires averaging across valuations in \mathcal{V} (normative uncertainty) and outcomes (empirical uncertainty). Yet our options effectively need just one dimension of averaging, as they have just one source of uncertainty. The first

option has just normative uncertainty: it surely has outcome x , of value 4 or 0. The second option has just empirical uncertainty: it has outcome y of sure value 4 or outcome z of sure value 0. Each option thus has the same average value of the outcome: $\frac{1}{2}4 + \frac{1}{2}0 = 2$.

- *Ex-Ante* Expectationalism forms the average value of the ex-ante value prospect, which is the same prospect $4_{50\%}0_{50\%}$ for each option. So, regardless of which option we consider, we must calculate $\frac{1}{2}v(4_{50\%}0_{50\%}) + \frac{1}{2}v'(4_{50\%}0_{50\%})$. What are $v(4_{50\%}0_{50\%})$ and $v'(4_{50\%}0_{50\%})$? As $4_{50\%}0_{50\%}$ is option 2's value prospect given v , $v(4_{50\%}0_{50\%}) = v(\text{option 2}) = 1$. As $4_{50\%}0_{50\%}$ is also option 2's value prospect given v' , $v'(4_{50\%}0_{50\%}) = v'(\text{option 2}) = 1$. So, $\frac{1}{2}v(4_{50\%}0_{50\%}) + \frac{1}{2}v'(4_{50\%}0_{50\%}) = \frac{1}{2}1 + \frac{1}{2}1 = 1$.
- *Reverse* Expectationalism forms the average value of the empirically ex-post value prospect. As for Ex-Post Expectationalism, this in principle requires averaging across both outcomes and valuations; but for each option one dimension of averaging drops out, because option 1 is empirically riskless and option 2 is normatively riskless. Option 1 surely has outcome x , whose value prospect $4_{50\%}0_{50\%}$ is evaluated at 1 by both (risk-averse) valuations, as just seen. The average value is thus $\frac{1}{2}1 + \frac{1}{2}1 = 1$. Option 2 either has outcome y , whose value prospect $4_{100\%}$ has value 4 under both v and v' ; or has outcome z , whose value prospect $0_{100\%}$ has value 0 under both v and v' . The average value is thus $\frac{1}{2}4 + \frac{1}{2}0 = 2$.

What attitudes to risk do the four expectational meta-theories take? Very different ones. We shall explain this point only informally here, since a separate paper focuses on risk attitudes (Dietrich and Jabarian 2021b). We construe risk aversion and other risk attitudes as attitudes to the risk *about outcome value*, such as (in our example) the risk of ending up either with value 4 or with value 0. This is not the only possible notion of risk attitudes. Different notions of risk attitudes differ in the quantity with respect to which risk is defined. For economists in the Arrow-Pratt tradition, this quantity is monetary wealth, or consumption, or some other empirical quantity; here, risk (aversion) lies in (aversion to) unknown wealth, or unknown consumption, etc. For other scholars, including ourselves, the quantity is outcome value; here, risk (aversion) consists in (aversion to) unknown outcome *value*.¹³ In general, being risk-averse (-prone, -neutral) with respect to a given quantity Q means that any option leading to a risky amount of Q is ranked below (above, like) receiving the option's *expected* amount of Q for sure; in the special case of von-Neumann-Morgenstern preferences, an equivalent condition is that vNM utility is convex (concave, linear) in Q .¹⁴

¹³ An example in theoretical economics is Grant et al. (2010, 2012), who indeed assume that an option constitutes the same risk for two individuals if it generates for each individual the same subjective value prospect, rather than wealth prospect or consumption prospect.

¹⁴ For instance, if outcomes are wealth levels, so that $X \subseteq \mathbb{R}$ and A contains 'wealth lotteries', and if risk is measured in wealth itself, then risk aversion means that any risky wealth lottery $a \in A$ is worse than getting the expected wealth $\bar{a} = \sum_{x \in X} a(x)x$ for sure; which is equivalent concavity of vNM utility, in the special case of von-Neumann-Morgenstern preferences. If instead, as we assume, risk aversion is defined w.r.t. *the value* of the outcome (here: of wealth), then the 'expectation test' is performed, not with wealth levels, but with their values. This notion of risk keeps risk aversion distinct from the phenomenon of diminishing marginal value of wealth. For instance, most individuals would prefer having an astronomic amount of wealth x for sure to facing a wealth lottery that yields wealth 0 or wealth $2x$ equiprobably – *because* the decreasing marginal value of wealth makes the difference

Risk about outcome value can have an empirical, or normative, or mixed origin: it can stem from uncertainty about the outcome (empirical uncertainty), or uncertainty about the value of the outcome (normative uncertainty), or uncertainty about both. A meta-theory is risk-averse if its evaluation of options contains a penalty for risk in outcome value, i.e., lies below the *expected* outcome value; it is risk-neutral if evaluations match the expected outcome value. One may want the meta-theory to ‘borrow’ the risk attitude of those first-order valuations that you find credible, i.e., have non-zero credence in. For instance, if you have positive credence only in risk-averse valuations in \mathcal{V} , then the meta-theory is risk-averse. We call this *risk-impartiality*, because your meta-level risk attitude defers to your risk-attitudinal judgments (Dietrich and Jabarian 2021b). We set aside what risk-impartiality requires when you are risk-attitudinally undecided, e.g., have non-zero credence both in a risk-averse valuation and a risk-neutral valuation. In our example, you are certain that risk-aversion is correct, as v and v' are both risk-averse; here a risk-impartial meta-theory is risk-averse.¹⁵ Defending the principle of risk-impartiality goes beyond this paper. We merely note, in Table 3, that our four meta-theories differ in their risk attitudes: one is risk-neutral no matter what your risk-attitudinal judgments are (no penalty for risk), one is risk-impartial (deference to risk-attitudinal judgments), and two have hybrid risk-attitudes, i.e., are risk-neutral or risk-impartial depending on the origin of risk. To explain why, we use

	neutral to normative risk	impartial to normative risk
neutral to empirical risk	Ex-Post Expectationalism	Reverse Expectationalism
impartial to empirical risk	Standard Expectationalism	Ex-Ante Expectationalism

Table 3: The risk attitudes of the four meta-theories in our example with risk-averse first-order theories

again our example, in which all credible valuations (i.e., v and v') are risk-averse, so that risk impartiality reduces to risk aversion.

- *Standard* Expectationalism applies the valuations v and v' to the option, which captures only empirical risk. This leads (by risk-aversion of v and v') to a penalty or discount for empirical risk only: the theory is averse to empirical risk, but neutral to normative risk. This explains why in Table 2 the normatively risky option 1 gets the undiscounted value of 2, while the empirically risky option 2 gets the discounted value of 1.
- *Ex-Post* Expectationalism applies the two valuations to the outcome, which captures no risk. So no risk is penalized: the theory is globally risk-neutral. This explains why both options in Table 2 get the undiscounted value of 2.
- *Ex-Ante* Expectationalism applies the two valuations to the ex-ante value prospect, which captures risk of both origins. So all risk is penalized: the theory is globally risk-averse. This explains why both options in Table 2 get the discounted

in value between x and $2x$ small compared to that between 0 and x . If risk attitudes were defined with respect to wealth rather than value of wealth, then this preference would be incompatible with risk-neutrality or risk-proneness.

¹⁵Risk attitudes have been analysed extensively in the different context of purely empirical uncertainty. For different accounts, see Weirich (1986), Buchak (2013), Bradley and Stefánsson (2017) and Baccelli (2018).

value of 1.

- *Reverse* Expectationalism applies the two valuations to the empirically ex-post value prospect, which captures only normative risk. So only normative risk is penalized: the theory is averse only to normative risk. This explains why in Table 2 only the normatively risky option gets the discounted value of 1.

6 The full taxonomy of Expectationalism

We now turn to a unification. We introduce a single generic expectational theory, of which our four earlier theories are nothing but special cases. The generic theory depends on a parameter that determines the reasoning mode, i.e., the extent of ex-post-ness. Particular choices of this parameter yield our four special expectational theories, and all other expectational theories. So there exist not just four expectational theories, but a large and unified class of expectational theories.

The parameter determining the expectational theory is the type of *information* relative to which reasoning is ex-post: full information yields Ex-Post Expectationalism, no information yields Ex-Ante Expectationalism, purely normative information yields Standard Expectationalism, purely empirical information yields Reverse Expectationalism, and yet other types of information yield other expectational theories.

We model an information by an *empirical-normative event* $I \subseteq X \times \mathcal{V}$, containing the empirical-normative worlds (x, v) which are consistent with the information. Salient examples:

- The full information of a specific empirical-normative world (x, v) is $I = \{(x, v)\}$, a singleton set of worlds.
- The vacuous or tautological information is $I = X \times \mathcal{V}$, the set of *all* worlds.
- The information of a specific valuation v is $I = X \times \{v\}$, the set of worlds of type $(*, v)$.
- The information of a specific outcome x is $I = \{x\} \times \mathcal{V}$, the set of worlds of type $(x, *)$.

Recall that each option a generates a world prospect, i.e., a probability function over worlds. Let us denote it by P_a . The probability of a world (x, v) is $P_a(x, v) = a(x)Pr(v)$, the product of the probabilities of x and v .

To define our general expectational theory, we need a general notion of value prospect. A general value prospect is based on a general degree of ex-post-ness: it conditionalises on a general information. Formally, for any option $a \in X$ and information $I \subseteq X \times \mathcal{V}$ (of non-zero probability $P_a(I)$), the *value prospect of a given I* is the value prospect $p_{a|I}$ such that the probability of a value level $k \in \mathbb{R}$ is the probability that a results in value k given I :

$$\begin{aligned} p_{a|I}(k) &= \text{probability of final value } k \text{ given } I = \frac{\text{prob. of } [I \text{ \& final value } k]}{\text{prob. of } I} \\ &= \frac{P_a(\{(x, v) \in I : v(x) = k\})}{P_a(I)}. \end{aligned}$$

Our four earlier notions of value prospect are special cases, obtained for special inform-

ation:

Proposition 1 *The value prospect $p_{a|I}$ of an option $a \in A$ given an information $I \subseteq X \times \mathcal{V}$ (of non-zero probability $P_a(I)$) coincides with the*

- *ex-ante value prospect p_a if $I = X \times \mathcal{V}$ (no information),*
- *ex-post value prospect $p_{x,v}$ if $I = \{(x, v)\}$ (information of a full world (x, v)),*
- *normatively ex-post value prospect $p_{a,v}$ if $I = X \times \{v\}$ (information of a valuation v),*
- *empirically ex-post value prospect p_x if $I = \{x\} \times \mathcal{V}$ (information of an empirical outcome x).*

Recall that each valuation v in \mathcal{V} can evaluate not just options, but (derivatively) also value prospects. So we can form $v(p_{a|I})$, which tells how valuable v finds the prospect of option a given I . We can call $v(p_{a|I})$ *a's value given I* or *ex- I value*, according to v .

An expectational theory reasons ex-post w.r.t. some type of information. A type of information is represented by an *information partition*: a partition \mathcal{I} of the set $X \times \mathcal{V}$ of empirical-normative worlds. \mathcal{I} contains those information I on which the reasoner conditionalises when conceptualizing options as prospects. As such, \mathcal{I} defines a degree of ex-post-ness of reasoning. Each reasoning mode draws on some partition \mathcal{I} . The four salient special cases are:

- Fully ex-post reasoning draws on the finest information partition, $\mathcal{I} = \{\{(x, v)\} : (x, v) \in X \times \mathcal{V}\}$.
- Fully ex-ante reasoning draws on the coarsest partition, $\mathcal{I} = \{X \times \mathcal{V}\}$.
- Normatively ex-post reasoning draws on the partition into ‘valuation events’, $\mathcal{I} = \{X \times \{v\} : v \in \mathcal{V}\}$.
- Empirically ex-post reasoning draws on the partition into ‘outcome events’, $\mathcal{I} = \{\{x\} \times \mathcal{V} : x \in X\}$.

Each information partition \mathcal{I} – each degree of ex-post reasoning – determines an expectational theory, which evaluates options by the expected value (across empirical-normative worlds (x, v)) of the prospect given \mathcal{I} . Formally:

Ex- \mathcal{I} Expectationalism (*EV \mathcal{I}*): *The meta-value of an option $a \in A$ is the expected value of the prospect given \mathcal{I} or expected ex- \mathcal{I} value.*¹⁶

$$EV_{\mathcal{I}}(a) = \sum_{(x,v) \in X \times \mathcal{V}} \underbrace{a(x)Pr(v)}_{\text{prob. of } (x,v)} v(p_{a|\mathcal{I}(x,v)})$$

where $\mathcal{I}(x, v)$ is the information in empirical-normative world (x, v) , i.e., the $I \in \mathcal{I}$ containing (x, v) .

We can now define ‘Expectationalism’ as a general approach or type of theory:

¹⁶ Although $p_{a|\mathcal{I}(x,v)}$ becomes undefined in the zero-probability case $P_a(\mathcal{I}(x, v)) = 0$, no ambiguity arises. Whenever $p_{a|\mathcal{I}(x,v)}$ is undefined, the value $v(p_{a|\mathcal{I}(x,v)})$ can be interpreted arbitrarily, as it is multiplied by 0 ($= P_a(x, v) = a(x)Pr(v)$) and so has no effect.

Expectationalism: *Meta-value is given by some expectational theory, i.e., by Ex- \mathcal{I} Expectationalism for some information type \mathcal{I} (some partition \mathcal{I} of $X \times \mathcal{V}$).*

Our four earlier theories are special cases, obtained using particular information types, i.e., by presupposing particular degrees of ex-post reasoning:

Theorem 2 *Ex- \mathcal{I} Expectationalism coincides with*

- *Ex-Ante Expectationalism if $\mathcal{I} = \{X \times \mathcal{V}\}$ (no information),*
- *Ex-Post Expectationalism if $\mathcal{I} = \{(x, v) : (x, v) \in X \times \mathcal{V}\}$ (full information),*
- *Standard Expectationalism if $\mathcal{I} = \{X \times \{v\} : v \in \mathcal{V}\}$ (normative information),*
- *Reverse Expectationalism if $\mathcal{I} = \{\{x\} \times \mathcal{V} : x \in X\}$ (empirical information).*

Are there any circumstances under which it becomes irrelevant how you reason? That is, can it happen that all degrees of ex-post reasoning yield the same expectational theory, hence the same evaluation of options, albeit through different procedures? This question obviously matters. If all reasoning modes were extensionally equivalent, then you could reason as you wish or find easiest. The question has a sharp answer:

Theorem 3 *All expectational theories $EV_{\mathcal{I}}$ coincide (i.e., your reasoning mode has no effect) if and only if you are certain that value is of vNM type, i.e., $Pr(v) = 0$ for all non-vNM-type valuations v in \mathcal{V} .*

Recall that a valuation is of vNM-type if the value of any option is the expected value of its outcome (see Section 1). Some scholars have argued that correct value is of vNM type, and many modellers routinely assume value is of vNM type for technical simplicity. But few people (if anyone at all) will be utterly certain that correct value is of vNM type. These few people can safely reason at any level of ex-post-ness they wish: their reasoning mode does not affect the resulting judgments by Theorem 3. But anyone who entertains at least some doubt that value is of vNM type faces the hard choice between expectational theories, hence between reasoning modes.

7 Choice under normative uncertainty as choice by Harsanyi's impartial observer or Rawls's person behind the veil of ignorance

Choice under normative uncertainty bears a clear analogy to choice by Harsanyi's (1955, 1978) 'impartial observer' or Rawls's (1971) person in the 'original position' behind a 'veil of ignorance', insofar as Harsanyi's and Rawls's imaginary agents do not know their own identity, hence their values. In fact, Harsanyi himself refers to "moral uncertainty" (Harsanyi 1975: 602). While Rawls argues for *maximin* choice under such ignorance, Harsanyi defends *expected-utility* choice, in apparent analogy to Standard Expectationalism under normative uncertainty.

But what exactly is the relationship between Harsanyi's theory and normative uncertainty? Like us, Harsanyi considers a choice under risk, given by a set of lotteries

A over an underlying set of outcomes X (technically, he requires X to be finite and A to contain *all* lotteries over X , two assumptions we make in this section). He assumes this choice is faced by a society of individuals $i = 1, 2, \dots, n$. Each individual, as well as society as a whole, holds a preference relation on A , denoted $\succeq_1, \dots, \succeq_n$ and \succeq , respectively. Harsanyi assumes that these preference relations satisfy von-Neumann-Morgenstern's three axioms (of 'weak order', 'continuity', and 'independence'). Hence, they are representable by vNM-type utility functions, denoted u_1, \dots, u_n and U , respectively. We here use Harsanyi's term "utility function" rather than "valuation", although both terms stand for real-valued function on A . Recall that such a function is said to be of 'vNM type' just in case it assigns to any lottery in A the expectation of what it assigns to the outcomes in X . By Harsanyi's seminal Aggregation Theorem, if, in addition to the vNM assumption, a standard Pareto condition is imposed (to relate social to individual preferences), then social vNM utility must be a linear combination of individual vNM utility:

$$U(a) = \sum_{i=1}^n w_i u_i(a) \text{ for all } a \in A, \quad (1)$$

for some fixed weights $w_i \geq 0$ ($i = 1, \dots, n$). This formula superficially resembles Standard Expectationalism. The two differences are (i) the use of vNM utility functions u_i ($i = 1, \dots, n$) instead of arbitrary valuations v ($\in \mathcal{V}$) and (ii) the use of arbitrary weights w_i instead of correctness probabilities $Pr(v)$.

Harsanyi derives the linear formula (1) a second time, using his (equally seminal) Impartial Observer Theorem. In this theorem, the aggregate vNM utility function U has a different interpretation: it represents the preferences of a hypothetical 'impartial observer' who does not know his identity. He could be any of the individuals in society, with known probabilities. Under the theorem's assumptions, the observer adopts preferences that combine those of the individuals, following the same linear formula (1) as in Harsanyi's Aggregation Theorem. But an individual i 's weight w_i now gets a new meaning: it is the probability with which the observer takes on i 's identity (and values). This interpretation of weights as identity probabilities tightens the link to (standard) expectational choice under normative uncertainty, formally and interpretationally. To be more precise (for interested readers), the impartial observer's preferences are initially defined on lotteries over *extended* outcomes, i.e., pairs (x, i) of an outcome in X and an individual in $\{1, \dots, n\}$, representing the situation of facing x as person i . A lottery over extended outcomes represents how likely it is to face such-and-such outcomes as such-and-such persons. But, after fixing the 'identity probabilities', i.e., postulating that any identity i is taken with some given probability w_i , the observer's preferences over extended lotteries induce ordinary lottery preferences.¹⁷ These ordinary lottery preferences capture the impartial observer's preferences *given* the identity probabilities

¹⁷How? Call \hat{A} the set of lotteries over the set $X \times \{1, \dots, n\}$ of extended outcomes. Fix for each individual i a probability w_i of being i , where the w_i 's are non-negative with sum 1. Then any ordinary lottery $a \in A$ can be 'extended' to one in \hat{A} (the probability of any extended outcome (x, i) being $a(x)w_i$, the product of the probabilities of x and i); and the observer's initial preference relation on \hat{A} induces a preference relation on A , according to which a lottery $a \in A$ is weakly preferred to another $b \in A$ if and only if a 's extension is weakly preferred to b 's extension.

w_i ($i = 1, \dots, n$). They are representable by the linear function (1), under the theorem’s assumptions.¹⁸

But does Harsanyi’s linear aggregation formula (1) really correspond to Standard Expectationalism? It does not, because the individuals’ vNM utility functions u_i need not be their utility functions in the relevant cardinal and interpersonally comparable sense. The distinction between vNM utility and ‘true’ utility is of course controversial; it is the object of the Harsanyi-Sen debate. For Harsanyi, vNM utility *is* utility in the proper sense. We side with Sen (and Weymark), who object that the vNM function u_i is just one of infinitely many possible numerical representations of i ’s value *order*; any increasing transformation of u_i would represent the same order. Except in a lucky coincidence, u_i does not measure i ’s utility in the full, non-ordinal and interpersonally comparable, sense. If each individual i has a ‘true’ utility function, denoted v_i , then the correct analogue of Standard Expectationalism is not a linear aggregation of vNM utility, given by (1), but a linear aggregation of ‘true’ utility, given by

$$V(a) = \sum_{i=1}^n w_i v_i(a) \text{ for all } a \in A. \quad (2)$$

Grant et al. (2010) have proved a version of Harsanyi’s Impartial Observer Theorem in which the observer’s preferences obey a more general representation than (1), obtained from (1) by replacing each individual vNM utility function u_i by some (continuous) increasing transformation of it.¹⁹ These transformed functions could be the ‘true’ utility functions v_i ; then Grant et al.’s observer preferences are given by (2), i.e., are analogous to standard expectational preferences under normative uncertainty.²⁰

8 Conclusion

There is more than one expected-value theory for evaluating options. The various expectational theories differ *ethically*, by reaching different evaluations; *procedurally*, by using different reasoning; and *risk-attitudinally*, by taking different attitudes towards empirical as well as normative risk. But all theories take the expected value of some

¹⁸Harsanyi is particularly interested in the “impartial” case in which all w_i ’s are identical, so that (1) becomes an unweighted linear average of individual utilities. This special case does not matter to us here, since we do not focus on social ethics, but on the relationship to normative uncertainty.

¹⁹Formally, each u_i is replaced by $\phi_i \circ u_i$ for some increasing function $\phi_i : \mathbb{R} \rightarrow \mathbb{R}$.

²⁰Grant et al. show that their observer preferences respond better to the risk attitudes of individuals than Harsanyi’s vNM-based observer preferences (1), which run into paradoxical counterexamples as soon as individuals have different risk attitudes. Grant et al.’s insight could be translated from his choice problem (choice by an impartial observer) to ours (choice under normative uncertainty). In a nutshell, the analogous implication is that Standard Expectationalism responds better to the risk attitudes of valuations in \mathcal{V} than an alternative meta-theory that takes the expectation, not of the valuations in \mathcal{V} , but of vNM functions ordinally equivalent to them (this meta-theory is only definable if such vNM functions exist, i.e., if the order induced by any v in \mathcal{V} obeys von-Neumann-Morgenstern’s axioms). In the terminology of Section 5, Standard Expectationalism is more properly risk-impartial than its vNM-based alternative. However, the risk impartiality of Standard Expectationalism is restricted to empirical risk, since Standard Expectationalism is notoriously risk-*neutral* with respect to normative risk (cf. Section 5).

type of prospect. At the two ends of the spectrum, *Ex-Ante* and *Ex-Post* Expectationalism respectively take the expected value of the ex-ante or ex-post prospect, hence reason from the perspective before or after resolution of any uncertainty (empirical or normative). *Standard* Expectationalism lies in between: it takes the expected value of the option itself, thereby effectively reasoning from an empirically ex-ante, but normatively ex-post perspective. *Reverse* Expectationalism does the opposite: it reasons empirically ex-post, but normatively ex-ante. The four mentioned theories stand out as salient, but they are just examples. In general, to any type of information (technically, to any ‘information partition’ of the set of empirical-normative worlds) corresponds an expectational theory, which reasons ex-post relative to *this* information.

The classical question ‘Expectationalism or not?’ should therefore be complemented by another pressing question: ‘Expected value of what?’ The problem of deciding between versions of Expectationalism might prove to be as difficult as the classic problem of deciding between ex-ante and ex-post versions of egalitarianism – or perhaps even more difficult, as we face normative uncertainty besides empirical uncertainty.

References

- Baccelli, J. (2018) Risk Attitudes in Axiomatic Decision Theory—A Conceptual Perspective, *Theory and Decision* 84: 61-82
- 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., Jabarian, B. (2021a) Axiomatic foundations of normative uncertainty, working paper
- Dietrich, F., Jabarian, B. (2021b) The risk attitude under normative uncertainty, working paper
- 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., Mongin, P. (2016) The utilitarian relevance of the aggregation theorem, *American Economic Journal: Microeconomics* 8(3): 289–306
- 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
- Grant S., Kajii, A., Polak, B., Safra, Z.. (2010) Generalized utilitarianism and Harsanyi's impartial observer theorem, *Econometrica* 78(6): 1939-71
- Grant S., Kajii, A., Polak, B., Safra, Z. (2012) Equally-distributed equivalent utility, ex post egalitarianism and utilitarianism, *Journal of Economic Theory* 147(4): 1545-1571
- Greaves, H., Cotton-Barratt, O. (2019) A bargaining-theoretic approach to moral uncertainty', working paper
- Greaves, H. (2017) A reconsideration of the Harsanyi-Sen-Weymark debate on utilitarianism, *Utilitas* 29(2): 175-213
- Hansen, L. P., Sargent, T. J. (2001) Robust Control and Model Uncertainty, *American Economic Review* 91: 60-66
- Harsanyi, J. (1955) Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility, *J Political Economy* 63: 309-21
- Harsanyi, J. (1975) Review: Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory, *The American Political Science Review* 69(2): 594-606
- 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.
URL: <http://commonsenseatheism.com/wp-content/uploads/2014/03/MacAskill-Normative-Uncertainty.pdf>
- MacAskill, W. (2016) Normative Uncertainty as a Voting Problem, *Mind* 125: 967-1004
- MacAskill, W., Ord, T. (2020) Why maximize expected choice-worthiness? *Noûs* 54(2): 327-353
- 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
- McCarthy, D. (2015) Distributive equality, *Mind* 124: 1045-1109
- 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
- Podgorski, A. (2019) Normative uncertainty and the dependence problem, *Mind* 129(513): 43-70
- Rawls, J. (1971) *A Theory of Justice*, Cambridge, Mass.: Harvard University Press
- Riedener, S. (2020) An axiomatic approach to axiological uncertainty, *Philosophical*

Studies 177: 483-504

Ross, J. (2006) Rejecting ethical deflationism, *Ethics* CXVI: 742–768

Rowe, T., Voorhoeve, A. (2018) Egalitarianism under Severe Uncertainty, *Philos Public Aff* 46: 239-268

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

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

Tarsney, C. (2018a) Moral Uncertainty for Deontologists, *Ethical Theory and Moral Practice* 21: 505–520

Tarsney, C. (2018b) Intertheoretic Value Comparison: A Modest Proposal, *Journal of Moral Philosophy* 15: 324-344

Tarsney, C. (2019) Normative uncertainty and social choice, *Mind*

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

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

A Proofs

Proof of Theorem 1. Let $a \in X$. Firstly,

$$EV_{ante}(a) = \sum_{v \in \mathcal{V}} Pr(v)v(p_a) = \sum_{v \in \mathcal{V}} Pr(v)v(p_a) \sum_{x \in X} a(x) = \sum_{(x,v) \in X \times \mathcal{V}} a(x)Pr(v)v(p_a),$$

where the second equality holds as $\sum_{x \in X} a(x) = 1$. Secondly,

$$\begin{aligned} EV_{stan}(a) &= \sum_{v \in \mathcal{V}} Pr(v)v(a) = \sum_{v \in \mathcal{V}} Pr(v)v(p_{a,v}) \\ &= \sum_{v \in \mathcal{V}} Pr(v)v(p_{a,v}) \sum_{x \in X} a(x) = \sum_{(x,v) \in X \times \mathcal{V}} a(x)Pr(v)v(p_{a,v}), \end{aligned}$$

where the second equality holds because $v(p_{a,x}) = v(a)$, and the third because $\sum_{x \in X} a(x) = 1$. Thirdly, the expression for $EV_{rev}(a)$ holds by definition. Finally,

$$EV_{post}(a) = \sum_{(x,v) \in X \times \mathcal{V}} a(x)Pr(v) \underbrace{v(x)}_{=v(p_{x,v})} = \sum_{(x,v) \in X \times \mathcal{V}} a(x)Pr(v)v(p_{x,v}). \blacksquare$$

Proof of Proposition 1. Consider an option $a \in A$ and an information $I \subseteq X \times \mathcal{V}$ such that $P_a(I) \neq 0$. As our definitions easily imply, if $I = X \times \mathcal{V}$ then $p_{a|I} = p_a$, while if $I = \{(x, v)\}$ where $(x, v) \in X \times \mathcal{V}$ then $p_{a|I} = p_{x,v}$. If $I = X \times \{v\}$ where $v \in \mathcal{V}$, then

$p_{a|I} = p_{a,v}$ because for all $k \in \mathbb{R}$

$$\begin{aligned} p_{a|I}(k) &= \frac{P_a(\{x \in X : v(x) = k\} \times \{v\})}{Pr(v)} = \frac{a(\{x \in X : v(x) = k\})Pr(v)}{Pr(v)} \\ &= a(\{x \in X : v(x) = k\}) = \sum_{x \in X: v(x)=k} a(x) = p_{a,v}(k). \end{aligned}$$

Finally, if $I = \{x\} \times \mathcal{V}$ where $x \in X$, then $p_{a|I} = p_x$ because for all $k \in \mathbb{R}$

$$\begin{aligned} p_{a|I}(k) &= \frac{P_a(\{x\} \times \{v \in \mathcal{V} : v(x) = k\})}{a(x)} = \frac{a(x)Pr(\{v \in \mathcal{V} : v(x) = k\})}{a(x)} \\ &= Pr(\{v \in \mathcal{V} : v(x) = k\}) = \sum_{v \in \mathcal{V}: v(x)=k} Pr(v) = p_x(k). \blacksquare \end{aligned}$$

Proof of Theorem 2. Regarding EV_{ante} , for each option $a \in A$

$$EV_{ante}(a) = \sum_{(x,v) \in X \times \mathcal{V}} a(x)Pr(v)v(p_a) = EV_{\mathcal{I}}(a) \text{ for } \mathcal{I} = \{X \times \mathcal{V}\}.$$

where the first identity holds by Theorem 1 and the second identity holds because by Proposition 1 we can replace p_a by $p_{a, X \times \mathcal{V}} = p_{a|\mathcal{I}(x,v)}$. Analogously, for each $a \in A$

$$\begin{aligned} EV_{stan}(a) &= \sum_{(x,v) \in X \times \mathcal{V}} a(x)Pr(v)v(\underbrace{p_a}_{p_{a, X \times \{v\}}}) = EV_{\mathcal{I}}(a) \text{ for } \mathcal{I} = \{X \times \{v\} : v \in \mathcal{V}\} \\ EV_{rev}(a) &= \sum_{(x,v) \in X \times \mathcal{V}} a(x)Pr(v)v(\underbrace{p_x}_{p_{a, \{x\} \times \mathcal{V}}}) = EV_{\mathcal{I}}(a) \text{ for } \mathcal{I} = \{\{x\} \times \mathcal{V} : x \in X\} \\ EV_{post}(a) &= \sum_{(x,v) \in X \times \mathcal{V}} a(x)Pr(v)v(\underbrace{p_{x,v}}_{p_{a, \{(x,v)\}}}) = EV_{\mathcal{I}}(a) \text{ for } \mathcal{I} = \{\{(x,v)\} : (x,v) \in X \times \mathcal{V}\}, \end{aligned}$$

where on each line the two identities use Theorem 1 and Proposition 1, respectively. \blacksquare

The proof of Theorem 3 begins with a lemma.

Lemma 1 *A valuation $v \in \mathcal{V}$ is of vNM type if and only if it evaluates value prospects by their expectation, i.e., $v(p) = Exp(p)$ ($= \sum_{k \in \mathbb{R}} p(k)k$) for all value prospects p .*

Proof. 1. First, let $v \in \mathcal{V}$ be of vNM type. We fix a value prospect p and prove that $v(p) = Exp(p)$. Pick an option $a \in A$ such that $p_{a,v} = p$. We have

$$Exp(p) = \sum_{k \in \mathbb{R}} kp(k) = \sum_{k \in \mathbb{R}} k \sum_{x \in X: v(x)=k} a(x) = \sum_{k \in \mathbb{R}} \sum_{x \in X: v(x)=k} a(x)k = \sum_{x \in X} a(x)v(x),$$

where the second equality uses that $p(k) = p_{a,v}(k) = \sum_{x \in X: v(x)=k} a(x)$, and the third and fourth equalities follow by reordering terms. The last expression equals $v(a)$ as v is of vNM-type, which equals $v(p)$ by choice of a .

2. Conversely, assume $v(p) = Exp(p)$ for all value prospects p . We let $a \in A$ and show $v(a) = \sum_{x \in X} a(x)v(x)$. Defining p as $p_{a,v}$, we have $Exp(p) = \sum_{x \in X} a(x)v(x)$, as in part 1 of the proof. So it remains to show $v(a) = Exp(p)$. This holds because $v(a) = v(p)$ (as $p = p_{a,v}$) and $v(p) = Exp(p)$ (by Q.e.d.sis). \blacksquare

Proof Theorem 3. We shall use standard measure-theoretic arguments.

1. Assume $Pr(v) = 0$ for all non-vNM-type valuations $v \in \mathcal{V}$. Fix an option $a \in A$. We show that $EV_{\mathcal{I}}(a)$ is independent of the information partition \mathcal{I} . On the set of worlds $X \times \mathcal{V}$, consider the probability distribution P_a (the world prospect of a) and the random variables $\mathbf{x} : X \times \mathcal{V} \rightarrow X$, $(x, v) \mapsto x$ and $\mathbf{v} : X \times \mathcal{V} \rightarrow \mathcal{V}$, $(x, v) \mapsto v$. Combining these variables yields a third variable, $\mathbf{v}(\mathbf{x})$, given by $X \times \mathcal{V} \rightarrow \mathbb{R}$, $(x, v) \mapsto v(x)$ and representing resulting value. The value prospect p_a equals the distribution of the variable $\mathbf{v}(\mathbf{x})$, and so its expectation is $Exp(p_a) = Exp_{P_a}(\mathbf{v}(\mathbf{x}))$. More generally, for any information $I \subseteq X \times \mathcal{V}$ (such that $P_a(I) \neq 0$), the value prospect $p_{a|I}$ equals the distribution of $\mathbf{v}(\mathbf{x})$ conditional on I , and so $Exp(p_{a|I}) = Exp_{P_a}(\mathbf{v}(\mathbf{x})|I)$. Now for any information partition \mathcal{I} (identifiable with the variable mapping (x, v) to $\mathcal{I}(x, v)$),

$$\begin{aligned} EV_{\mathcal{I}}(a) &= Exp_{P_a}(\mathbf{v}(p_{a|\mathcal{I}})) && \text{by definition} \\ &= Exp_{P_a}(Exp(p_{a|\mathcal{I}})) && \text{by Lemma 1} \\ &= Exp_{P_a}(Exp_{P_a}(\mathbf{v}(\mathbf{x})|\mathcal{I})) && \text{as } Exp(p_{a|\mathcal{I}}) = Exp_{P_a}(\mathbf{v}(\mathbf{x})|\mathcal{I}) \\ &= Exp_{P_a}(\mathbf{v}(\mathbf{x})) && \text{by the law of iterated expectations,} \end{aligned}$$

where Lemma 1 is applicable as valuations generated by \mathbf{v} (with non-zero probability) are of vNM type. The last expression for $EV_{\mathcal{I}}(a)$ shows that $EV_{\mathcal{I}}(a)$ is independent of \mathcal{I} .

2. Conversely, let \mathcal{V} contain a non-vNM-type valuation \tilde{v} of probability $Pr(\tilde{v}) \neq 0$. As \tilde{v} is not of vNM type, we may pick an option $a \in A$ such that $\tilde{v}(a) \neq \sum_{x \in X} a(x)\tilde{v}(x)$. Denote the information of valuation \tilde{v} by $I = X \times \{\tilde{v}\}$. We construct two information partitions \mathcal{I}_1 and \mathcal{I}_2 for which $EV_{\mathcal{I}_1}(a) \neq EV_{\mathcal{I}_2}(a)$. Let \mathcal{I}_1 and \mathcal{I}_2 coincide outside I and be, respectively, maximally coarse or maximally fine within I . So $\mathcal{I}_1 = \mathcal{I}_0 \cup \{I\}$ and $\mathcal{I}_2 = \mathcal{I}_0 \cup \{(x, v) : (x, v) \in I\}$, for some partition \mathcal{I}_0 of $(X \times \mathcal{V}) \setminus I$. Thus $EV_{\mathcal{I}_1}(a) = S + S_1$ and $EV_{\mathcal{I}_2}(a) = S + S_2$ where

$$\begin{aligned} S &= \sum_{(x,v) \in (X \times \mathcal{V}) \setminus I} a(x)Pr(v)v(p_{a|\mathcal{I}_0(x,v)}) \\ S_1 &= \sum_{(x,v) \in I} a(x)Pr(v)v(p_{a|\mathcal{I}_1(x,v)}) = \sum_{x \in X} a(x)Pr(\tilde{v})\tilde{v}(p_{a|I}) \\ S_2 &= \sum_{(x,v) \in I} a(x)Pr(v)v(p_{a|\mathcal{I}_2(x,v)}) = \sum_{x \in X} a(x)Pr(\tilde{v})\tilde{v}(p_{a,\{(x,\tilde{v})\}}). \end{aligned}$$

By Proposition 1, $p_{a|I} = p_{a,\tilde{v}}$ and $p_{a,\{(x,\tilde{v})\}} = p_{x,\tilde{v}}$. So $\tilde{v}(p_{a|I}) = \tilde{v}(p_{a,\tilde{v}}) = \tilde{v}(a)$ and $\tilde{v}(p_{a,\{(x,\tilde{v})\}}) = \tilde{v}(p_{x,\tilde{v}}) = \tilde{v}(x)$. Thus

$$\begin{aligned} S_1 &= \sum_{x \in X} a(x)Pr(\tilde{v})\tilde{v}(a) = Pr(\tilde{v})\tilde{v}(a) \sum_{x \in X} a(x) = Pr(\tilde{v})\tilde{v}(a) \\ S_2 &= \sum_{x \in X} a(x)Pr(\tilde{v})\tilde{v}(x) = Pr(\tilde{v}) \sum_{x \in X} a(x)\tilde{v}(x). \end{aligned}$$

So

$$EV_{\mathcal{I}_1}(a) - EV_{\mathcal{I}_2}(a) = S_1 - S_2 = Pr(\tilde{v}) \left(\tilde{v}(a) - \sum_{x \in X} a(x)\tilde{v}(x) \right).$$

As $Pr(\tilde{v}) \neq 0$ and $\tilde{v}(a) \neq \sum_{x \in X} a(x)\tilde{v}(x)$, we deduce $EV_{\mathcal{I}_1}(a) \neq EV_{\mathcal{I}_2}(a)$. ■