SYMMETRY OF MODELS VERSUS MODELS OF SYMMETRY

GERT DE COOMAN AND ENRIQUE MIRANDA

ABSTRACT. A model for a subject's beliefs about a phenomenon may exhibit symmetry, in the sense that it is invariant under certain transformations. On the other hand, such a belief model may be intended to represent that the subject believes or knows that the phenomenon under study exhibits symmetry. We defend the view that these are fundamentally different things, even though the difference cannot be captured by Bayesian belief models. In fact, the failure to distinguish between both situations leads to Laplace's so-called Principle of Insufficient Reason, which has been criticised extensively in the literature.

We show that there are belief models (imprecise probability models, coherent lower previsions) that generalise and include the Bayesian belief models, but where this fundamental difference can be captured. This leads to two notions of symmetry for such belief models: weak invariance (representing symmetry of beliefs) and strong invariance (modelling beliefs of symmetry). We discuss various mathematical as well as more philosophical aspects of these notions. We also discuss a few examples to show the relevance of our findings both to probabilistic modelling and to statistical inference, and to the notion of exchangeability in particular.

1. INTRODUCTION

This paper deals with symmetry in relation to models of beliefs. Consider a model for a subject's beliefs about a certain phenomenon. Such a *belief model* may be *symmetrical*, in the sense that it is invariant under certain transformations. On the other hand, a belief model may try to capture that the subject believes that the phenomenon under study exhibits symmetry, and we then say that the belief model *models symmetry*. We defend the view that there is an important conceptual difference between the two cases: symmetry of beliefs should not be confused with beliefs of symmetry.[1]

Does this view need defending at all? That there is a difference may strike you as obvious, and yet we shall argue that Bayesian belief models, which are certainly the most popular belief models in the literature, are unable to capture this difference.

To make this clearer, consider a simple example. Suppose I will toss a coin, and you are ignorant about its relevant properties: it might be fair but on the other hand it might be heavily loaded, or it might even have two heads, or two tails (situation *A*). To you the outcomes of the toss that are practically possible are *h* (for heads) and *t* (for tails). Since you are ignorant about the properties of the coin, any model for your beliefs should not change if heads and tails are permuted, so the model that 'faithfully' captures your beliefs about the outcome of the toss should be symmetrical too, i.e., invariant under this permutation of heads and tails.

---

[1]This echoes Walley's [1991, Section 9.5.6, p. 466] view that 'symmetry of evidence' is not the same thing as 'evidence of symmetry'.

Suppose on the other hand that you know that the coin (and the tossing mechanism) I shall use is completely symmetrical (situation *B*). Your belief model about the outcome of the toss should capture this knowledge, i.e., it should model your beliefs about the symmetry of the coin.

Our point is that belief models should be able to catch the important difference between your beliefs in the two situations. Bayesian belief models cannot do this. Indeed—the argument is well-known—the only symmetrical probability model, which is in other words invariant under permutations of heads and tails, assigns equal probability 1/2 to heads and tails. But this is automatically also the model that captures your beliefs that the coin is actually symmetrical, so heads and tails should be equally likely.

The real reason why Bayesian belief models cannot capture the difference between symmetry of models and modelling symmetry, is that they do not allow for *indecision*. Suppose that I ask you to express your preferences between two gambles, whose reward depends on the outcome of the toss. For first one, *a*, you will win one euro if the outcome is heads, and lose one if it is tails. The second one, *b*, gives the same rewards, but with heads and tails swapped.

In situation *B*, because you believe the coin to be symmetrical, it does not matter to you which gamble you get, and you are *indifferent* in your choice between the two.

But in situation *A*, on the other hand, because you are completely ignorant about the coin, the available information gives you *no reason to (strictly) prefer a* over *b* or *b* over *a*. You are therefore *undecided* about which of the two gambles to choose.

Because decision based on Bayesian belief models leaves you no alternative but to either strictly prefer one action over the other, or to be indifferent between them, the symmetry of the model leaves you *no choice but to act as if you were indifferent* between *a* and *b*. We strongly believe that it is wrong to confuse indecision with indifference in this example (and elsewhere of course), but Bayesian belief models leave you no choice but to do so, unless you want to let go of the principle that if your evidence or your beliefs are symmetrical, your belief model should be symmetrical as well. The problem with Laplace's Principle of Insufficient Reason is precisely this: if you use a Bayesian probability model then the symmetry present in ignorance forces you to treat indecision (or insufficient reason to decide) between *a* and *b* as if it were indifference.[2] Or in other words, it forces you to treat symmetry of beliefs as if there were beliefs of symmetry.

If on the other hand, we consider belief models that allow for indecision, we can sever the unholy link between indecision and indifference, because in a state of complete ignorance, we are then allowed to remain undecided about which of the two actions to choose: in the language of preference relations, they simply become *incomparable*, and you need not be indifferent between them. As we shall see further on, similar arguments show that such belief models also allow us to distinguish between 'symmetry of models' and 'models of symmetry' in those more general situations where the symmetry involved is not necessarily that which goes along with complete ignorance.

So, it appears that in order to better understand the interplay between modelling beliefs and issues of symmetry, which is the main aim of this paper, we shall need to work with a language, or indeed, with a type of belief models that, unlike the Bayesian ones, take

---

[2]This may seem a good explanation why Keynes [1921, p. 83] renamed the 'Principle of Insufficient Reason' the 'Principle of Indifference'. He (and others, see Zabell [1989b]) also suggested that the principle should not be applied in a state of complete ignorance, but only if there is good reason to justify the indifference (such as when there is evidence of symmetry). By the way, Keynes was also among the first to consider what we shall call imprecise probability models, as his comparative probability relations were not required to be complete.

indecision seriously. For this purpose, we shall use the language of the so-called *imprecise probability models* [Walley, 1991], and in particular coherent lower previsions, which have the same behavioural pedigree as the more common Bayesian belief models (*in casu* coherent previsions, see de Finetti [1974–1975]), and which contain these models as a special case. We give a somewhat unusual introduction to such models in Section 2.[3] In Section 3, we provide the necessary mathematical background for discussing symmetry: we discuss monoids of transformations, and invariance under such monoids. After these introductory sections, we start addressing the issue of symmetry in relation to belief models in Section 4. We introduce two notions of invariance for the imprecise probability models introduced in Section 2: *weak invariance*, which captures symmetry of belief models, and *strong invariance*, which captures that a model represents the belief that there is symmetry. We study relevant mathematical properties of these invariance notions, and argue that the distinction between them is very relevant when dealing with symmetry in general, and in particular (Section 5) for modelling complete ignorance. Further interesting properties of weak and strong invariance, related to inference, are the subject of Sections 6 and 7, respectively. We show among other things that a weakly invariant coherent lower prevision can always be extended to a larger domain, in a way that is as conservative as possible. This implies that, for any given monoid of transformations, there always are weakly invariant coherent lower previsions. This is not generally the case for strong invariance, however, and we give and discuss sufficient conditions such that for a given monoid of transformations, there would be strongly invariant coherent (lower) previsions. We also give various expression for the smallest strongly invariant coherent lower prevision that dominates a given weakly invariant one (if it exists). In Section 8, we turn to the important example of coherent (lower) previsions on the set of natural numbers, that are shift-invariant, and we use them to characterise the strongly invariant coherent (lower) previsions on a general space provided with a single transformation. Further examples are discussed in Section 9, where we characterise weak and strong invariance with respect to finite groups of permutations. In particular, we discuss Walley's [1991] generalisation to lower previsions of de Finetti's [1937] notion of exchangeability, and we use our characterisation of strong permutation invariance to prove a generalisation to lower previsions of de Finetti's representation results for finite sequences of exchangeable random variables. Conclusions are gathered in Section 10.

We want to make it clear at this point that this paper owes a significant intellectual debt to Peter Walley. First of all, we use his behavioural imprecise probability models [Walley, 1991] to try and clarify the distinction between symmetry of beliefs and beliefs of symmetry. Moreover, although we like to believe that much of what we do here is new, we are also aware that in many cases we take to their logical conclusion a number of ideas about symmetry that are clearly present in his work (mainly Walley [1991, Sections 3.5, 9.4 and 9.5] and Pericchi and Walley [1991]), sometimes in embryonic form, and often more fully worked out.

2. Imprecise probability models

Consider a very general situation in which uncertainty occurs: a subject is uncertain about the value that a variable X assumes in a set of possible values $\mathscr{X}$. Because the subject is uncertain, we shall call X an *uncertain*, or *random*, variable.

---

[3]For other brief and perhaps more conventional introductions to the topic, we refer to Walley [1996a], De Cooman and Zaffalon [2004], De Cooman and Troffaes [2004], De Cooman and Miranda [2006]. A much more detailed account of the behavioural theory of imprecise probabilities can be found in Walley [1991].

The central concept we shall use in order to model our subject's uncertainty about X, is that of a *gamble* (on X, or on $\mathscr{X}$), which is a bounded real-valued function f on $\mathscr{X}$. In other words, a gamble f is a map from $\mathscr{X}$ to the set of real numbers $\mathbb{R}$ such that

$$\sup f := \sup\{f(x)\colon x \in \mathscr{X}\} \text{ and } \inf f := \inf\{f(x)\colon x \in \mathscr{X}\}$$

are (finite) real numbers. It is interpreted as the reward function for a transaction which may yield a different (and possibly negative) reward $f(x)$, measured in units (called *utiles*) of a pre-determined linear utility,[4] for each of the different values x that the random variable X may assume in $\mathscr{X}$.

We denote the set of all gambles on X by $\mathscr{L}(\mathscr{X})$. For any two gambles f and g, we denote their point-wise sum by $f + g$, and we denote the point-wise (scalar) multiplication of f with a real number λ by λf. $\mathscr{L}(\mathscr{X})$ is a real linear space under these operations. We shall always endow this space with the *supremum norm*, i.e., $\|f\| = \sup|f| = \sup\{|f(x)|\colon x \in \mathscr{X}\}$, or equivalently, with the topology of uniform convergence, which turns $\mathscr{L}(\mathscr{X})$ into a Banach space.

An *event A* is a subset of $\mathscr{X}$. If $X \in A$ then we say that the event *occurs*, and if $X \notin A$ then we say that *A doesn't occur*, or equivalently, that the *complement(ary event) $A^c = \{x \in \mathscr{X} : x \notin A\}$* occurs. We shall identify an event with a special $\{0,1\}$-valued gamble I_A, called its *indicator*, and defined by $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ elsewhere. We shall often write A for I_A, whenever there is no possibility of confusion.

2.1. **Coherent sets of really desirable gambles.** Given the information that the subject has about X, she will be disposed to accept certain gambles, and to reject others. The idea is that we model a subject's beliefs about X by looking at which gambles she accepts, and to collect these into a *set of really desirable gambles $\mathscr{R}$*.

The dice example. Assume that our subject is uncertain about the outcome X of my tossing a die. In this case $\mathscr{X} = \mathscr{X}_6 := \{1,2,3,4,5,6\}$ is the set of possible values for X. If the subject is rational, she will accept the gamble which yields a positive reward whatever the value of X, because she is certain to improve her 'fortune' by doing so. On the other hand, she will not accept a non-positive gamble that is negative somewhere, because by accepting such a gamble she can only lose utility (we then say she *incurs a partial loss*). She will not accept the gamble which makes her win one utile if the outcome X is 1, and makes her lose five utiles otherwise, unless she knows for instance that the die is loaded very heavily in such a way that the outcome 1 is almost certain to come up.

Real desirability can also be interpreted in terms of the betting behaviour of our subject. Suppose she wants to bet on the occurrence of some event, such as my throwing 1 (so that she receives 1 utile if the event happens and 0 utiles otherwise). If she thinks that the die is fair, she should be disposed to bet on this event at any rate r strictly smaller than $\frac{1}{6}$. This means that the gamble $I_{\{1\}} - r$ representing this transaction (winning $1 - r$ if the outcome of X is 1 and losing r otherwise) will be really desirable to her for $r < \frac{1}{6}$. ♦

Now, accepting certain gambles has certain consequences, and has certain implications for accepting other gambles, and if our subject is rational, which we shall assume her to be, she should take these consequences and implications into account. To give but one example, if our subject accepts a certain gamble f she should also accept any other gamble

---

[4]This utility can be regarded as amounts of money, as is the case for instance in de Finetti [1974–1975]. It is perhaps more realistic, in the sense that the linearity of the scale is better justified, to interpret it in terms of probability currency: we win or lose lottery tickets depending on the outcome of the gamble; see Walley [1991, Section 2.2].

g such that $g \geq f$, i.e., such that g *point-wise dominates* f, because accepting g is certain to bring her a reward that is at least as high as accepting f does.

Actually, this requirement is a consequence [combine (D2) with (D3)] of the following four basic rationality axioms for real desirability, which we shall assume any rational subject's set of really desirable gambles $\mathscr{R}$ to satisfy:

(D1) if $f < 0$ then $f \notin \mathscr{R}$ [avoiding partial loss];

(D2) if $f \geq 0$ then $f \in \mathscr{R}$ [accepting sure gains];

(D3) if $f \in \mathscr{R}$ and $g \in \mathscr{R}$ then $f + g \in \mathscr{R}$ [accepting combined gambles]

(D4) if $f \in \mathscr{R}$ and $\lambda > 0$ then $\lambda f \in \mathscr{R}$ [scale invariance].

where $f < g$ is shorthand for $f \leq g$ and $f \neq g$.[5] We call any subset $\mathscr{R}$ of $\mathscr{L}(\mathscr{X})$ that satisfies these axioms a *coherent* set of really desirable gambles.

It is easy to see that these axioms reflect the behavioural rationality of our subject: (D1) means that she should not be disposed to accept a gamble which makes her lose utiles, no matter the outcome; (D2) means that she should accept a gamble which never makes her lose utiles; on the other hand, if she is disposed to accept two gambles f and g, she should also accept the combination of the two gambles, which leads to a reward $f + g$; this is an immediate consequence of the linearity of the utility scale. This justifies (D3). And finally, if she is disposed to accept a gamble f, she should be disposed to accept the scaled gamble λf for any $\lambda > 0$, because this just reflects a change in the linear utility scale. This is the idea behind condition (D4).

Walley [1991, 2000] has a further coherence axiom that sets of really desirable gambles should satisfy, which turns out to be quite important for conditioning, namely

(D5) if $\mathscr{B}$ is a partition of $\mathscr{X}$ and if $I_B f \in \mathscr{R}$ for all B in $\mathscr{B}$, then $f \in \mathscr{R}$ [full conglomerability].

Since this axiom is automatically satisfied whenever $\mathscr{X}$ is finite [it is then an immediate consequence of (D3)], and since we shall not be concerned with conditioning unless when $\mathscr{X}$ is finite (see Section 9), we shall ignore this additional axiom in the present discussion.

A coherent set of really desirable gambles is a convex cone [axioms (D3)–(D4)] that includes the 'non-negative orthant' $\mathscr{C}_+ := \{f \in \mathscr{L}(\mathscr{X}) : f \geq 0\}$ [axiom (D2)] and has no gamble in common with the 'negative orthant' $\mathscr{C}_- := \{f \in \mathscr{L}(\mathscr{X}) : f < 0\}$ [axiom (D1)].[6] If we have two coherent sets of really desirable gambles $\mathscr{R}_1$ and $\mathscr{R}_2$, such that $\mathscr{R}_1 \subseteq \mathscr{R}_2$, then we say that $\mathscr{R}_1$ is less committal, or more conservative, than $\mathscr{R}_2$, because a subject whose set of really desirable gambles is $\mathscr{R}_2$ accepts at least all the gambles in $\mathscr{R}_1$. The least-committal (most conservative, smallest) coherent set of really desirable gambles is $\mathscr{C}_+$. Within this theory, it seems to be the appropriate model for *complete ignorance*: if our subject has no information at all about the value of X, she should be disposed to accept only those gambles which cannot lead to a loss of utiles (see also the discussion in Section 5).

Now suppose that our subject has specified a set $\mathscr{R}$ of gambles that she accepts. In an elicitation procedure, for instance, this would typically be a finite set of gambles, so we cannot expect this set to be coherent. We are then faced with the problem of enlarging this $\mathscr{R}$ to a coherent set of really desirable gambles that is as small as possible: we want to find out what are the (behavioural) consequences of the subject's accepting the gambles in $\mathscr{R}$, taking into account *only* the requirements of coherence. This inference problem is

---

[5]So, here and in what follows, we shall write '$f < 0$' to mean '$f \leq 0$ and not $f = 0$', and '$f > 0$' to mean '$f \geq 0$ and not $f = 0$'.

[6]This means that the zero gamble 0 belongs to the set of really desirable gambles. This is more a mathematical convention than a behavioural requirement, since this gamble has no effect whatsoever in the amount of utiles of our subject. See more details in Walley [1991].

(also formally) similar to the problem of inference (logical closure) in classical proposi-
tional logic, where we want to find out what are the consequences of accepting certain
propositions.[7]

The smallest convex cone including $\mathscr{C}_+$ and $\mathscr{R}$, or in other words, the smallest subset
of $\mathscr{L}(\mathscr{X})$ that includes $\mathscr{R}$ and satisfies (D2)–(D4), is given by

$$\mathscr{E}^r_{\mathscr{R}} := \left\{ g \in \mathscr{L}(\mathscr{X}) : g \geq \sum_{k=1}^n \lambda_k f_k \text{ for some } n \geq 0, \lambda_k \in \mathbb{R}^+ \text{ and } f_k \in \mathscr{R} \right\},$$

where $\mathbb{R}^+$ denotes the set of non-negative real numbers. If this convex cone $\mathscr{E}^r_{\mathscr{R}}$ intersects
$\mathscr{C}_-$ then it is easy to see that actually $\mathscr{E}^r_{\mathscr{R}} = \mathscr{L}(\mathscr{X})$, and then it is impossible to extend $\mathscr{R}$
to a coherent set of really desirable gambles [because (D1) cannot be satisfied]. Observe
that $\mathscr{E}^r_{\mathscr{R}} \cap \mathscr{C}'_- = \emptyset$ if and only if

$$\text{there are no } n \geq 0, \lambda_k \in \mathbb{R}^+ \text{ and } f_k \in \mathscr{R} \text{ such that } \sum_{k=1}^n \lambda_k f_k < 0,$$

and we then say that the set $\mathscr{R}$ *avoids partial loss*. Let us interpret this condition. As-
sume that it doesn't hold (so we say that $\mathscr{R}$ *incurs partial loss*). Then there are really
desirable gambles $f_1, \ldots, f_n$ and positive $\lambda_1, \ldots, \lambda_n$ such that $\sum_{k=1}^n \lambda_k f_k < 0$. But if our
subject is disposed to accept the gamble f_k then by coherence [axioms(D2) and (D4)] she
should also be disposed to accept the gamble $\lambda_k f_k$ for all $\lambda_k \geq 0$. Similarly, by coherence
[axiom (D3)] she should also be disposed to accept the sum $\sum_{k=1}^n \lambda_k f_k$. Since this sum is
non-positive, and strictly negative in at least some elements of $\mathscr{X}$, we see that the subject
can be made subject to a partial loss, by suitably combining gambles which she accepts.
This is unreasonable.

When the class $\mathscr{R}$ avoids partial loss, and only then, we are able to extend $\mathscr{R}$ to a
coherent set of really desirable gambles, and the smallest such set is precisely $\mathscr{E}^r_{\mathscr{R}}$, which is
called the *natural extension* of $\mathscr{R}$ to a set of really desirable gambles. This set reflects only
the behavioural consequences of the assessments present in $\mathscr{R}$: the acceptance of a gamble
f not in $\mathscr{E}^r_{\mathscr{R}}$ (or, equivalently, a set of really desirable gambles strictly including $\mathscr{E}^r_{\mathscr{R}}$) is not
implied by the information present in $\mathscr{R}$, and therefore represents stronger implications
that those of coherence alone.

2.2. **Coherent sets of almost-desirable gambles.** Coherent sets of really desirable gam-
bles constitute a very general and powerful class of models for a subject's beliefs (see
Walley [1991, Appendix F] and Walley [2000] for more details and discussion). We could
already discuss symmetry aspects for such coherent sets of really desirable gambles, but
we shall instead concentrate on a slightly less general and powerful type of belief models,
namely coherent lower and upper previsions. Our main reason for doing so is that this will
allow us to make a more direct comparison to the more familiar Bayesian belief models,
and in particular to de Finetti's [1974–1975] coherent previsions, or fair prices.

Consider a gamble f. Then our subject's *lower prevision*, or supremum acceptable
buying price, $\underline{P}(f)$ for f is defined as the largest real number s such that she accepts the
gamble $f - t$ for any price $t < s$, or in other words accepts to buy f for any such price t.
Similarly, her *upper prevision*, or infimum acceptable selling price, $\overline{P}(f)$ for the gamble f
is the smallest real number s such that she accepts the gamble $t - f$ for any price $t > s$, or
in other words accepts to sell f for any such price t.

---

[7]See Moral and Wilson [1995] and De Cooman [2000, 2005] for more details on this connection between
natural extension and inference in classical propositional logic.

For an event A, the lower prevision $\underline{P}(I_A)$ of its indicator is also called the *lower probability* of A, and denoted by $\underline{P}(A)$. It can be interpreted as the supremum rate for betting on the event A. Similarly, $\overline{P}(I_A)$ is called the *upper probability* of A, and also denoted by $\overline{P}(A)$.

Since selling a gamble f for price s is the same thing as buying $-f$ for price $-s$, we have the following *conjugacy* relationship between an upper and a lower prevision:

$$\overline{P}(f) = -\underline{P}(-f).$$

This implies that from a given lower prevision $\underline{P}$, we can always construct the conjugate upper prevision $\overline{P}$, so they are mathematically equivalent belief models. In what follows, we shall mainly concentrate on lower previsions.

Now assume that our subject has a coherent set of really desirable gambles $\mathscr{R}$, then it is clear from the definition of lower and upper prevision that we can use $\mathscr{R}$ to define a lower prevision

$$\underline{P}_{\mathscr{R}}(f) = \sup\{s \in \mathbb{R} : f - s \in \mathscr{R}\} \tag{D-LPR}$$

and an upper prevision

$$\overline{P}_{\mathscr{R}}(f) = \inf\{s \in \mathbb{R} : s - f \in \mathscr{R}\}$$

for every gamble f on $\mathscr{X}$. So, given $\mathscr{R}$ we can construct two real-valued functionals, $\underline{P}_{\mathscr{R}}$ and $\overline{P}_{\mathscr{R}}$, whose interpretation is that of a supremum acceptable buying price, and an infimum acceptable selling price, respectively, and whose domain is $\mathscr{L}(\mathscr{X})$. We shall call these functionals *lower* and *upper previsions*.

We call a *coherent lower prevision* on $\mathscr{L}(\mathscr{X})$ any real-valued functional on $\mathscr{L}(\mathscr{X})$ satisfying the following three axioms:

(P1) $\underline{P}(f) \geq \inf f$ [accepting sure gains];
(P2) $\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$ [super-additivity];
(P3) $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ [non-negative homogeneity].

for all gambles f and g on $\mathscr{X}$, and all non-negative real λ.

It follows from the coherence axioms (D1)–(D4) for $\mathscr{R}$ that the lower prevision $\underline{P}_{\mathscr{R}}$ that corresponds to a coherent set of really desirable gambles $\mathscr{R}$ is coherent.[8]

So we see that with a coherent set of really desirable gambles $\mathscr{R}$, we can define a coherent lower prevision on $\mathscr{L}(\mathscr{X})$, using (D-LPR). We shall see further on that, conversely, given a coherent lower prevision $\underline{P}$ on $\mathscr{L}(\mathscr{X})$, we can always find a coherent set of really desirable gambles $\mathscr{R}$ such that $\underline{P}$ and $\mathscr{R}$ are related through (D-LPR). But unfortunately, the relationship between the two types of belief models is many-to-one: there are usually many coherent sets of really desirable gambles that lead to the same coherent lower prevision. This is why we said before that coherent sets of really desirable gambles are a more general and powerful belief model than coherent lower previsions. The ultimate reason for this is the following: suppose that a subject specifies her supremum buying price $\underline{P}(f)$ for a gamble f. This implies that she accepts all the gambles $f - \underline{P}(f) + \delta$, where $\delta > 0$. But the specification of $\underline{P}(f)$ says nothing about the gamble $f - \underline{P}(f)$ (where $\delta = 0$) itself: she might accept it, but then again she might not. And precisely because specifying a coherent lower prevision says nothing about this border behaviour, it leads to a belief model that is less powerful than coherent sets of really desirable gambles, where this border behaviour would be determined.

The dice example (cont.). Let us go back to the die example. Consider, for any x in $\mathscr{X}_6 = \{1,\dots,6\}$, the event $\{x\}$ that the outcome X of rolling the die is x. If, for some real

---

[8]To prove (P1), use (D2); for (P2) use (D3); and for (P3) use (D4) for $\lambda > 0$ and (D1) and (D2) for $\lambda = 0$.

number r, our subject accepts the gamble $I_{\{x\}} - r$, she is willing to pay r utiles in return for the uncertain reward $I_{\{1\}}$, or in other words to bet *on* the event $\{1\}$ at *rate r*. So her lower probability $\underline{P}(\{x\})$ for $\{x\}$, or equivalently, her lower prevision $\underline{P}(I_{\{x\}})$ for $I_{\{x\}}$, is the supremum rate at which she is willing to bet on $\{x\}$. This means that she accepts the gamble $I_{\{x\}} - s$ for any $s < \underline{P}(\{x\})$. But it doesn't imply that she actually accepts the gamble $I_{\{x\}} - \underline{P}(\{x\})$: this gamble is only claimed to be almost-desirable, as we shall see further on.

If she is completely ignorant about the properties of the die, her evidence about the die is symmetrical, i.e., doesn't change when the possible outcomes are permuted. A belief model that 'faithfully' captures the available evidence should therefore be symmetrical with respect to such permutations as well, so we infer that in particular $\underline{P}(\{1\})$, ..., $\underline{P}(\{6\})$ are all equal to some number p. Coherence [use (P1) and (P2)] then requires that $0 \le p \le \frac{1}{6}$. Any such p leads to a symmetrical lower probability defined on the singletons, and therefore reflects 'symmetry of beliefs'. As we have indicated above, the model corresponding to $p = 0$ is the one that reflects complete ignorance. We shall see further on (see Sections 4.2 and 9) that the choice $p = \frac{1}{6}$ leads to the only model that captures the belief that the die is fair, i.e., that reflects 'beliefs of symmetry'. ♦

In order to better understand the relationship between coherent lower previsions and coherent sets of really desirable gambles, we need to introduce, besides *real* desirability, an new and weaker notion, called *almost-desirability*, which will also play an important part in our discussion of symmetry further on. This notion is inspired by the ideas in the discussion above: we say that a gamble f is *almost-desirable* to a subject, or that she *almost-accepts* f, whenever she accepts $f + \delta$, or in other words $f + \delta$ is really desirable to her, for any strictly positive amount of utility $\delta > 0$. By stating that f is almost-desirable to her, nothing is specified about whether the subject accepts f itself: she might, but then again she also might not. If we generically denote by $\mathscr{D}$ a set of gambles that are almost-desirable to our subject, we see that the set $\mathscr{D}_{\mathscr{R}}$ of almost-desirable gambles that corresponds to a coherent set $\mathscr{R}$ of really desirable gambles, is given by

$$\mathscr{D}_{\mathscr{R}} = \{f \in \mathscr{L}(\mathscr{X}) \colon (\forall \delta > 0) f + \delta \in \mathscr{R}\} = \bigcap_{\delta > 0} [\mathscr{R} - \delta] \qquad \text{(D-M)}$$

so $\mathscr{D}_{\mathscr{R}}$ is the closure (in the topology of uniform convergence on $\mathscr{L}(\mathscr{X})$) of the convex cone $\mathscr{R}$.

We call any set of gambles $\mathscr{D}$ that satisfies the following five axioms a *coherent set of almost-desirable gambles*:

(M1) if $\sup f < 0$ then $f \notin \mathscr{D}$ [avoiding sure loss];
(M2) if $\inf f \ge 0$ then $f \in \mathscr{D}$ [accepting sure gains];
(M3) if $f \in \mathscr{D}$ and $g \in \mathscr{D}$ then $f + g \in \mathscr{D}$ [accepting combined gambles];
(M4) if $f \in \mathscr{D}$ and $\lambda > 0$ then $\lambda f \in \mathscr{D}$ [scale invariance];
(M5) if $f + \delta \in \mathscr{D}$ for all $\delta > 0$ then $f \in \mathscr{D}$ [closure].

It is a closed and convex cone in $\mathscr{L}(\mathscr{X})$ that includes the non-negative orthant $\mathscr{C}_+$ and does not intersect with the set $\mathscr{C}'_- = \{f \in \mathscr{L}(\mathscr{X}) \colon \sup f < 0\} \subset \mathscr{C}_-$. It is easy to see that the set of almost-desirable gambles $\mathscr{D}_{\mathscr{R}}$ that corresponds to a coherent set of really desirable gambles $\mathscr{R}$ is actually also coherent.[9]

---

[9] To prove (M1), use (D1) with $\delta = -\frac{\sup f}{2}$; to prove (M2), use (D2); to prove (M3), use (D3); to prove (M4), use (D4); and to prove (M5), use $\varepsilon = \frac{\delta}{2}$ and the definition of $\mathscr{D}_{\mathscr{R}}$ to prove that $f + \delta \in \mathscr{R}$ for all $\delta > 0$.

It should at this point come as no surprise that coherent lower previsions and coherent sets of almost-desirable gambles are actually equivalent belief models. Indeed, consider a coherent set of almost-desirable gambles $\mathscr{D}$, i.e., $\mathscr{D}$ satisfies (M1)–(M5). Then the real-valued functional $\underline{P}_{\mathscr{D}}$ defined on $\mathscr{L}(\mathscr{X})$ by[10]

$$\underline{P}_{\mathscr{D}}(f) := \max\{s \in \mathbb{R} \colon f - s \in \mathscr{D}\} \qquad \text{(M-LPR)}$$

satisfies (P1)–(P3) and therefore is a coherent lower prevision on $\mathscr{L}(\mathscr{X})$.[11]

Conversely, if we consider a coherent lower prevision $\underline{P}$ on $\mathscr{L}(\mathscr{X})$, i.e., $\underline{P}$ satisfies (P1)–(P3), then the set of gambles

$$\mathscr{D}_{\underline{P}} := \{f \in \mathscr{L}(\mathscr{X}) \colon \underline{P}(f) \geq 0\} \qquad \text{(LPR-M)}$$

satisfies (M1)–(M5) and is therefore a coherent set of almost-desirable gambles.[12] Moreover, the relationships (M-LPR) and (LPR-M) are bijective (one-to-one and onto), and they are each other's inverses.[13]

Finally, consider a coherent lower prevision $\underline{P}$ on $\mathscr{L}(\mathscr{X})$, and define the following set of gambles

$$\mathscr{D}_{\underline{P}}^{+} := \{f \in \mathscr{L}(\mathscr{X}) \colon \underline{P}(f) > 0 \text{ or } f > 0\}.$$

Then $\mathscr{D}_{\underline{P}}^{+} \cup \{0\}$ is a coherent set of really desirable gambles, i.e., it satisfies (D1)–(D4).[14] Moreover, any coherent set of really desirable gambles $\mathscr{R}$ that satisfies

$$\mathscr{D}_{\underline{P}}^{+} \cup \{0\} \subseteq \mathscr{R} \subseteq \mathscr{D}_{\underline{P}},$$

i.e., the union of whose (relative) topological interior with $\mathscr{C}^{+}$ is $\mathscr{D}_{\underline{P}}^{+} \cup \{0\}$ and whose topological closure is $\mathscr{D}_{\underline{P}}$, has $\underline{P}$ as its associated lower prevision, through (D-LPR). This confirms what we claimed before: coherent lower previsions, or equivalently, coherent sets of almost-desirable gambles, are less powerful belief models than coherent sets of really desirable gambles. If a subject specifies a coherent lower prevision $\underline{P}$, then she actually states that all gambles in the union $\mathscr{D}_{\underline{P}}^{+} \cup \{0\}$ of $\mathscr{C}_{+}$ with the relative topological interior of $\mathscr{D}_{\underline{P}}$ are really desirable, but she doesn't specify whether the gambles in the topological boundary $\mathscr{D}_{\underline{P}} \setminus \mathscr{D}_{\underline{P}}^{+}$ of $\mathscr{D}_{\underline{P}}$ are: we only know that they are almost-desirable to her.

2.3. **Natural extension for coherent lower previsions.** There is one important problem that we skipped over in the discussion above, namely that of inference. Suppose a subject specifies a set $\mathscr{D}$ of gambles that are almost-desirable to her. In an elicitation procedure, for instance, this would typically be a finite set of gambles, so we cannot expect this set to be coherent. We are then, as before for really desirable gambles, faced with the problem of enlarging this $\mathscr{D}$ into a coherent set of almost-desirable gambles that is as small as possible: we want to find out what are the (behavioural) consequences of the subject's almost-accepting the gambles in $\mathscr{D}$, taking into account *only* the requirements of coherence.

---

[10]The supremum in Eq. (D-LPR) now becomes a maximum, simply because the set $\mathscr{D}$ is closed.

[11](P1) follows from (M2), (P2) from (M3) and (P3) is a consequence of (M4).

[12]First, conditions (P1) and (P2) imply that $\underline{P}$ is monotone. Now, (P2) and (P3) imply that $0 = \underline{P}(0) \geq \underline{P}(f) + \underline{P}(-f) \geq \underline{P}(f) + \inf(-f)$, whence $\underline{P}(f) \leq \sup f$. From these two facts we deduce (M1). (M2) is a consequence of (P1), (M3) of (P2) and (M4) of (P3). Finally, the monotonicity of $\underline{P}$ implies that $\underline{P}(\mu) = \mu$ for any constant value μ, and from this we deduce that $\underline{P}(f + \delta) = \underline{P}(f) + \delta$ for any $\delta > 0$. This implies (M5).

[13]To see that they are each other inverses, it suffices to use that a coherent lower prevision satisfies $\underline{P}(f - s) = \underline{P}(f) - s$ for any gamble f and any real number s, and, conversely, that $f \in \mathscr{D}_{\underline{P}}$ if and only if $\underline{P}(f) \geq 0$; this implies also that both transformations are bijective.

[14]For (D1), use that a coherent lower prevision $\underline{P}$ satisfies $\underline{P}(f) \leq \sup f$ for any gamble f; for (D2), that $f \geq 0$ satisfies either $f > 0$ or $f = 0$; for (D3), use (P2) and the monotonicity of the coherent $\underline{P}$, and for (D4) use (P3).

The smallest closed convex cone including $\mathscr{C}_+$ and $\mathscr{D}$, or in other words, the smallest subset of $\mathscr{L}(\mathscr{X})$ that includes $\mathscr{D}$ and satisfies (M2)–(M5), is given by

$$\mathscr{E}_{\mathscr{D}}^m := \left\{ g \in \mathscr{L}(\mathscr{X}) \colon (\forall \delta > 0)(\exists n \geq 0, \lambda_k \in \mathbb{R}^+, f_k \in \mathscr{D}) g \geq \sum_{k=1}^n \lambda_k f_k - \delta \right\}. \quad \text{(M-NE)}$$

This is the topological closure of the set $\mathscr{E}_{\mathscr{D}}^r$. If this convex cone $\mathscr{E}_{\mathscr{D}}^m$ intersects $\mathscr{C}_-^l = \{ f \in \mathscr{L}(\mathscr{X}) \colon \sup f < 0 \}$ then it is easy to see that actually $\mathscr{E}_{\mathscr{D}}^m = \mathscr{L}(\mathscr{X})$, and then it is impossible to extend $\mathscr{D}$ to a coherent set of almost-desirable gambles [because (M1) cannot be satisfied]. Observe that $\mathscr{E}_{\mathscr{D}}^m \cap \mathscr{C}_-^l = \emptyset$ if and only if[15]

$$\sup \left[\sum_{k=1}^n \lambda_k f_k \right] \geq 0 \text{ for some } n \geq 0, \ \lambda_k \in \mathbb{R}^+ \text{ and } f_k \in \mathscr{D}, \quad \text{(M-ASL)}$$

and we then say that the set $\mathscr{D}$ of almost-desirable gambles *avoids sure loss*. In that case, and only then, we are able to extend $\mathscr{D}$ to a coherent set of almost-desirable gambles, and the smallest such set is precisely $\mathscr{E}_{\mathscr{D}}^m$, which is called the *natural extension* of $\mathscr{D}$ to a set of almost-desirable gambles.

What does natural extension mean for the equivalent model of coherent lower previsions? Suppose our subject specifies a supremum acceptable buying price, or lower prevision, $\underline{P}(f)$ for each gamble f in some set of gambles $\mathscr{K} \subseteq \mathscr{L}(\mathscr{X})$.[16] We can then interpret $\underline{P}$ as a real-valued map on $\mathscr{K}$, and we call $\underline{P}$ a *lower prevision on* $\mathscr{K}$, and say that $\mathscr{K}$ is the *domain* of $\underline{P}$.

To study the problem of natural extension for this lower prevision, we shall use what we already know about natural extension in the context of almost-desirable gambles. Recall that specifying $\underline{P}$ on $\mathscr{K}$ is tantamount to stating that the gambles in the set $\mathscr{D} := \{ f - \underline{P}(f) \colon f \in \mathscr{K} \}$ are almost-desirable. We now look at the natural extension of this $\mathscr{D}$. Using (M-ASL), we know that such a natural extension exists if and only if[17]

$$\sup \left[\sum_{k=1}^n \lambda_k \left[f_k - \underline{P}(f_k) \right] \right] \geq 0 \text{ for all } n \geq 0, \ \lambda_k \in \mathbb{R}^+ \text{ and } f_k \in \mathscr{K}, \quad \text{(LPR-ASL)}$$

and we then say that the lower prevision $\underline{P}$ on $\mathscr{K}$ *avoids sure loss*. In this case, the natural extension $\mathscr{E}_{\mathscr{D}}^m$ is the smallest coherent set of almost-desirable gambles that includes $\mathscr{D}$, and consequently the coherent lower prevision $\underline{P}_{\mathscr{E}_{\mathscr{D}}^m}$ associated with $\mathscr{E}_{\mathscr{D}}^m$ through

$$\underline{P}_{\mathscr{E}_{\mathscr{D}}^m}(g) := \max \{ s \colon g - s \in \mathscr{E}_{\mathscr{D}}^m \}$$

is the point-wise smallest coherent lower prevision on $\mathscr{L}(\mathscr{X})$ that dominates $\underline{P}$ on $\mathscr{K}$. We call this coherent lower prevision the *natural extension* of $\underline{P}$ and we denote it by $\underline{E}_P$. We deduce from (M-NE) that for all gambles g on $\mathscr{X}$:

$$\underline{E}_P(g) = \sup_{\substack{\lambda_k \geq 0, g_k \in \mathscr{D} \\ k=1\ldots,n, n \geq 0}} \inf \left[g - \sum_{k=1}^n \lambda_k g_k \right] = \sup_{\substack{\lambda_k \geq 0, f_k \in \mathscr{K} \\ k=1\ldots,n, n \geq 0}} \inf \left[g - \sum_{k=1}^n \lambda_k \left[f_k - \underline{P}(f_k) \right] \right].$$
(LPR-NE)

If $\underline{P}$ *incurs sure loss*, i.e., (LPR-ASL) is not satisfied, then $\mathscr{E}_{\mathscr{D}}^m = \mathscr{L}(\mathscr{X})$ and consequently $\underline{E}_P$ assumes the value $+\infty$ in every gamble.

---

[15]Actually, this condition is equivalent to the one where we always choose $\lambda_k = 1$.

[16]This set of gambles $\mathscr{K}$ need not have any predefined structure; in particular, it does not have to be a linear space.

[17]Here too, this condition is equivalent to the one where we always choose $\lambda_k = 1$.

We shall call the lower prevision $\underline{P}$ on $\mathscr{K}$ *coherent*, whenever it can be extended to a coherent lower prevision on $\mathscr{L}(\mathscr{X})$, or in other words, whenever it coincides with its natural extension $\underline{E}_P$ on every gamble in its domain $\mathscr{K}$. Taking into account (LPR-NE), we see that this happens exactly when

$$\sup\left[\sum_{k=1}^{n} \lambda_k \left[f_k - \underline{P}(f_k)\right] - \lambda_0 \left[f_0 - \underline{P}(f_0)\right]\right] \geq 0 \text{ for all } n \geq 0,\, \lambda_k \in \mathbb{R}^+ \text{ and } f_k \in \mathscr{K},$$

(LPR-COH)

This coherence condition implies that $\underline{P}$ avoids sure loss.

Let us see if, for lower previsions, we can give a more immediate behavioural interpretation for avoiding sure loss, coherence, and natural extension. This should allow us to develop more intuition, as the approach we have followed so far, which motivates these notions through the coherence axioms for real and almost-desirable gambles, is admittedly quite abstract. We begin with avoiding sure loss. Suppose that condition (LPR-ASL) is not satisfied. Then there are $n \geq 0$, $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}^+$ and $f_1, \ldots f_n$ in $\mathscr{K}$ such that $\sup[\sum_{k=1}^{n} \lambda_k [f_k - \underline{P}(f_k)]] < 0$, which implies that there is some $\delta > 0$ for which

$$\sum_{k=1}^{n} \lambda_k \left[f_k - \underline{P}(f_k) + \delta\right] \leq -\delta.$$

Now, by the definition of $\underline{P}(f_k)$, our subject accepts each of the gambles $f_k - \underline{P}(f_k) + \delta$, so she should also accept the combined gamble $\sum_{k=1}^{n} \lambda_k[f_k - \underline{P}(f_k) + \delta]$ [use axioms (D3) and (D4) for real desirability]. But this gamble leads to a sure loss of at least δ. In other words, if condition (LPR-ASL) doesn't hold, there are gambles which the subject accepts and which, if properly combined, make her subject to a sure loss.

Next, assume that condition (LPR-COH) fails to hold. Then there are $n \geq 0$, $\lambda_0, \ldots, \lambda_n$ in $\mathbb{R}^+$ and $f_0, \ldots f_n$ in $\mathscr{K}$ such that $\sup[\sum_{k=1}^{n} \lambda_k[f_k - \underline{P}(f_k)] - \lambda_0[f_0 - \underline{P}(f_0)]] < 0$. Assume that $\lambda_0 > 0$, as we have already considered the case $\lambda_0 = 0$ in our discussion of avoiding sure loss. Then there is some $\delta > 0$ such that

$$\sum_{k=1}^{n} \frac{\lambda_k}{\lambda_0} \left[f_k - \underline{P}(f_k) + \delta\right] \leq f_0 - (\underline{P}(f_0) + \delta).$$

As before, the gamble on the left-hand side is a gamble that our subject accepts. But then she should also accept the gamble $f_0 - (\underline{P}(f_0) + \delta)$ since it point-wise dominates a gamble she accepts [use (D2) and (D3)]. This implies that she should be willing to pay a price $\underline{P}(f_0) + \delta$ for f_0, which is strictly higher than the supremum price $\underline{P}(f_0)$ she has specified for it. Coherence avoids this kind of inconsistency.

Finally, we turn to natural extension. Consider a gamble g on $\mathscr{X}$, then (LPR-NE) tells us that $\underline{E}_P(g)$ is the supremum s such that there are $n \geq 0$, $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}^+$ and $f_1, \ldots f_n$ in $\mathscr{K}$ for which

$$g - s \geq \sum_{k=1}^{n} \lambda_k \left[f_k - \underline{P}(f_k)\right]$$

Now the expression on the right-hand side is almost-desirable, because it is a non-negative linear combination of almost-desirable gambles [apply the axioms (M3) and (M4)]. So $g - s$ should be almost-desirable as well [apply the axioms (M2) and (M3)], and therefore our subject should be willing to buy g for any price $t < s$. So we deduce that $\underline{E}_P(g)$ is the supremum price for g that the subject can be forced to pay for the gamble g, by suitably combining transactions that she is committed to accept by her specifying the lower prevision $\underline{P}$ on $\mathscr{K}$. In other words, $\underline{E}_P(g)$ is the lower prevision for g that is implied by the assessments in $\underline{P}$ and coherence *alone*.

2.4. **Coherent previsions: the Bayesian belief models.** When a lower prevision $\underline{P}$ on $\mathscr{K}$ is *self-conjugate*, that is, when $\underline{P}(f) = \overline{P}(f)$ for any gamble f in $\mathscr{K}$, it is called a *prevision*. The common value $P(f)$ is then called the *prevision* of f; it is a *fair price* for the gamble f in the sense of de Finetti [1974–1975]. Formally, a real-valued function P on a class of gambles $\mathscr{K}$ is called a *linear*, or *coherent, prevision* whenever

$$\sup\left[\sum_{k=1}^{n}[f_k - P(f_k)] - \sum_{j=1}^{m}[g_j - P(g_j)]\right] \geq 0 \text{ for all } n,m \geq 0 \text{ and } f_k, g_j \in \mathscr{K}, \quad \text{(PR-COH)}$$

A linear prevision is coherent, both as a lower and as an upper prevision. Moreover, if its domain is the class of all gambles, $\mathscr{L}(\mathscr{X})$, then condition (PR-COH) simplifies to

(PR1) $P(f + g) = P(f) + P(g)$ for any f and g in $\mathscr{L}(\mathscr{X})$ [linearity].
(PR2) $P(f) \geq \inf f$ for any f in $\mathscr{L}(\mathscr{X})$ [accepting sure gains].

Linear previsions are the familiar Bayesian belief models: any linear prevision on all gambles is indeed a coherent prevision in the sense of de Finetti [1974–1975]; and a prevision defined on an arbitrary set of gambles is coherent exactly when it is the restriction of some coherent prevision on all gambles. The restriction to (indicators of) events of a coherent prevision on all gambles is a finitely additive probability. We shall denote by $\mathbb{P}(\mathscr{X})$ the set of all coherent previsions on $\mathscr{L}(\mathscr{X})$.

There is an interesting relationship between coherent previsions and coherent lower previsions. Let $\underline{P}$ be a lower prevision with domain $\mathscr{K}$, and let us denote by

$$\mathscr{M}(\underline{P}) := \{P \in \mathbb{P}(\mathscr{X}): (\forall f \in \mathscr{K})P(f) \geq \underline{P}(f)\}$$

the set of all coherent previsions on $\mathscr{L}(\mathscr{X})$ that *dominate* $\underline{P}$ on its domain. Then it can be checked[18] that $\underline{P}$ avoids sure loss if and only if $\mathscr{M}(\underline{P})$ is non-empty, that is, if and only if there is some coherent prevision on $\mathscr{L}(\mathscr{X})$ that dominates $\underline{P}$ on $\mathscr{K}$, and $\underline{P}$ is coherent if and only if it is the *lower envelope* of $\mathscr{M}(\underline{P})$, meaning that for all $\underline{P}$ in $\mathscr{K}$,

$$\underline{P}(f) = \min\{P(f): P \in \mathscr{M}(\underline{P})\}.$$

Also, any lower envelope of a set of coherent previsions is a coherent lower prevision. Moreover, the natural extension $\underline{E}_P$ of $\underline{P}$ to all gambles can be calculated using the set $\mathscr{M}(\underline{P})$ of coherent previsions: for any gamble f on $\mathscr{X}$, we have

$$\underline{E}_P(f) = \min\{P(f): P \in \mathscr{M}(\underline{P})\}.$$

This means that from a *mathematical* point of view, a coherent lower prevision $\underline{P}$ and its set of dominating coherent lower previsions $\mathscr{M}(\underline{P})$, are equivalent belief models. It can be checked that this set is convex and closed in the weak* topology.[19] Moreover, there is a bijective relationship between weak*-closed convex sets of coherent previsions and coherent lower previsions (their lower envelopes). This fact can (but need not) be used to give coherent lower previsions a *Bayesian sensitivity analysis interpretation*, besides the direct behavioural interpretation given in Section 2.2: we might assume the existence of a precise but unknown coherent prevision P expressing a subject's behavioural dispositions, and we might model the information about P by means of a weak*-closed convex set of coherent previsions $\mathscr{M}$ (the set of possible candidates). Then, this set is *mathematically* equivalent to its lower envelope $\underline{P}$, which is a coherent lower prevision. We shall come back to the difference between the direct behavioural and the Bayesian sensitivity analysis

---

[18]See [Walley, 1991, Sections 3.3–3.4] for proofs for these statements.

[19]The weak* topology on the set of all continuous linear functionals on $\mathscr{L}(\mathscr{X})$ is the topology of point-wise convergence. For more details, see Walley [1991, Appendix D].

interpretation of a lower prevision in Section 4.2, when we discuss the interplay between these interpretations and the notion of symmetry.

Taking into account the bijective relationship that exists between coherent lower previsions and sets of almost-desirable gambles, we may also establish a bijective relationship between sets of coherent previsions and sets of almost-desirable gambles: given a weak*-closed convex set $\mathscr{M}$ of coherent previsions on $\mathscr{L}(\mathscr{X})$, the class

$$\mathscr{D}_{\mathscr{M}} := \{f \in \mathscr{L}(\mathscr{X}): (\forall P \in \mathscr{M})P(f) \geq 0\}$$

is a coherent set of almost-desirable gambles, that is, it satisfies the coherence conditions (M1)–(M5). Conversely, given a coherent set of almost-desirable gambles $\mathscr{D}$, the corresponding set of coherent previsions

$$\mathscr{M}(\mathscr{D}) := \{P \in \mathbb{P}(\mathscr{L}): (\forall f \in \mathscr{D})P(f) \geq 0\}$$

is a weak*-closed convex set of coherent previsions.

Hence, there are at least three mathematically equivalent representations for the behavioural dispositions of our subject: coherent sets of almost-desirable gambles, coherent lower previsions, and weak*-closed convex sets of coherent previsions. The bijective relationships between them are summarised in Table 1.

↙	$\mathscr{D}$	$\underline{P}(\cdot)$	$\mathscr{M}$
$\mathscr{D}$		$\{f: \underline{P}(f) \geq 0\}$	$\{f: (\forall P \in \mathscr{M})P(f) \geq 0\}$
$\underline{P}(\cdot)$	$\max\{s: \cdot - s \in \mathscr{D}\}$		$\min\{P(\cdot): P \in \mathscr{M}\}$
$\mathscr{M}$	$\{P: (\forall f \in \mathscr{D})P(f) \geq 0\}$	$\{P: (\forall f)P(f) \geq \underline{P}(f)\}$	

TABLE 1. Bijective relationships between the equivalent belief models: coherent sets of almost-desirable gambles $\mathscr{D}$, coherent lower previsions $\underline{P}$ on $\mathscr{L}(\mathscr{X})$, and weak*-closed convex sets $\mathscr{M}$ of coherent previsions on $\mathscr{L}(\mathscr{X})$

We now briefly discuss a number of belief models that constitute particular instances of coherent lower previsions. First, we consider *n-monotone lower previsions*, where $n \geq 1$. A lower prevision $\underline{P}$ is called *n-monotone*[20] when the following inequality holds for all $p \in \mathbb{N}$, $p \leq n$, and all $f, f_1, \ldots, f_p$ in $\mathscr{L}(\mathscr{X})$:

$$\sum_{I \subseteq \{1,\ldots,p\}} (-1)^{|I|} \underline{P}\left(f \wedge \bigwedge_{i \in I} f_i\right) \geq 0,$$

where, here and further on, $|I|$ denotes the number of elements in a finite set I. A similar definition can be given if the domain of $\underline{P}$ is only a *lattice of gambles*, i.e., a set of gambles closed under point-wise minimum $\wedge$ and point-wise maximum $\vee$. Such *n-monotone lower previsions* are particular instances of exact functionals [Maaß, 2003], i.e., they are scalar multiples of some coherent lower prevision. In particular, an *n-monotone lower probability* defined on a lattice of events $\mathscr{S}$ that contains $\emptyset$ and $\mathscr{X}$ is coherent if and only if $\underline{P}(\emptyset) = 0$ and $\underline{P}(\mathscr{X}) = 1$.

---

[20]See De Cooman et al. [2006, 2005b,a] for a detailed discussion of *n*- and complete monotonicity for lower previsions.

A *completely monotone* lower prevision is simply one that is n-monotone for any natural number $n \geq 1$. When it is defined on indicators of events, it is called a completely monotone lower probability. When $\mathscr{X}$ is finite, this leads to *belief functions* in the terminology of Shafer [1976].

Two particular cases of belief functions and their conjugate upper probabilities are *probability charges*, or finitely additive probabilities defined on a field of events [Bhaskara Rao and Bhaskara Rao, 1983] and *possibility measures*. The latter [De Cooman, 2001, Zadeh, 1978] are set functions Π satisfying $\Pi\left(\bigcup_{i \in I} A_i\right) = \sup_{i \in I} \Pi(A_i)$ for any family $(A_i)_{i \in I}$ of subsets of $\mathscr{X}$. Π is a coherent *upper* probability if and only if $\Pi(\mathscr{X}) = 1$.

Finally, we can consider a particular instance of a completely monotone coherent lower prevision that allows us to model complete ignorance, the so-called *vacuous lower prevision*. It is given by

$$\underline{P}_{\mathscr{X}}(f) = \inf_{x \in \mathscr{X}} f(x),$$

for all gambles f on $\mathscr{X}$. It corresponds to the set of almost-desirable gambles $\mathscr{D} = \mathscr{C}_+ = \{f : f \geq 0\}$, and to the set $\mathscr{M} = \mathbb{P}(\mathscr{L})$ of all coherent previsions on $\mathscr{L}$. If we have no information at all about the values that X takes in $\mathscr{X}$, we have no reason to reject any coherent prevision P, and this leads to the vacuous lower prevision as a belief model. More generally, we can consider a vacuous lower prevision relative to some subset A of $\mathscr{X}$, which is given by

$$\underline{P}_A(f) = \inf_{x \in A} f(x).$$

A vacuous lower prevision relative to a set A is the adequate belief model when we know that the random variable X assumes values in A, and nothing else. The restriction to events of a vacuous upper prevision is a (zero-one-valued) possibility measure.

2.5. **Incomparability and indifference.** We claimed in the Introduction that Bayesian belief models do not take indecision seriously, and that we therefore need to look at a larger class of belief models that do not have this defect. Here, we present a better motivation for this claim.

Consider two gambles f and g on $\mathscr{X}$. We say that a subject *almost-prefers* f to g, and denote this as $f \succeq g$, whenever she accepts to exchange g for f in return for any (strictly) positive amount of utility. Given this definition, it is straightforward to check that we can express this in terms of the three equivalent belief models $\mathscr{D}$, $\underline{P}$ and $\mathscr{M}$ of the previous sections by

$$\begin{aligned}
f \succeq g &\Leftrightarrow f - g \in \mathscr{D} \\
&\Leftrightarrow \underline{P}(f - g) \geq 0 \\
&\Leftrightarrow (\forall P \in \mathscr{M}) P(f) \geq P(g).
\end{aligned}$$

The binary relation $\succeq$ is a partial pre-order on $\mathscr{L}(\mathscr{X})$, i.e., it is reflexive and transitive.[21] Observe also that $f \succeq g \Leftrightarrow f - g \succeq 0$ and that $f \succeq 0 \Leftrightarrow f \in \mathscr{D}$, so f is almost-preferred to g if and only if $f - g$ is almost-preferred to the zero gamble, which in turn is equivalent to the fact that our subject *almost-accepts* $f - g$, i.e., that $f - g$ is almost-desirable to her.

Unless our subject's lower prevision $\underline{P}$ is actually a (precise) prevision P (meaning that $\mathscr{D}$ is the semi-space $\{f : P(f) \geq 0\}$, and that $\mathscr{M} = \{P\}$), this ordering is not linear, or total: it does not hold for all gambles f and g that $f \succeq g$ or $g \succeq f$. When, therefore, both $f \not\succeq g$

---

[21]The binary relation $\succeq$ is actually a *vector ordering* on the linear space $\mathscr{L}(\mathscr{X})$, because it is compatible with the addition of gambles, and the scalar multiplication of gambles with non-negative real numbers.

and $g \nsucceq f$, we say that both gambles are *incomparable*, or that the subject is undecided about choosing between f and g, and we write this as $f \parallel g$.

It is instructive to see why the relation $\parallel$ is non-empty unless $\underline{P}$ is a precise prevision P. If $\underline{P}$ is not precise (but coherent), there is some gamble h such $\underline{P}(h) < \overline{P}(h)$. Let x be any real number such that $\underline{P}(h) < x < \overline{P}(h)$. In this case, the subject does not express a willingness to buy h for the price x, because x is strictly greater than her supremum acceptable price $\underline{P}(h)$ for buying h. Nor does she express a willingness to sell h for a price x, because x is strictly smaller than her infimum acceptable price $\overline{P}(h)$ for selling h. But there is more. Consider the gambles $f := h - x$ (buying h for a price x) and $g := x - h$ (selling h for a price x). Then it follows from the coherence of $\underline{P}$ that

$$\underline{P}(f-g) = 2\underline{P}(h-x) = 2[\underline{P}(h) - x] < 0 \text{ and } \underline{P}(g-f) = 2\underline{P}(x-h) = 2[x - \overline{P}(h)] < 0,$$

so $f \parallel g$: our subject is also undecided in the choice between buying h for x or selling h for that price.

We say that our subject is *indifferent* between f and g, and denote this as $f \approx g$ whenever both $f \succeq g$ and $g \succeq f$. This means that $\underline{P}(f-g) = \underline{P}(g-f) = 0$, or equivalently, $P(f) = P(g)$ for all P in $\mathcal{M}$. Clearly, $\approx$ is an equivalence relation (a reflexive, symmetrical and transitive binary relation) on $\mathscr{L}(\mathscr{X})$. It is important to distinguish between incomparability and indifference. Indifference between gambles f and g represents strong behavioural dispositions: it means that our subject almost-accepts to exchange f for g and *vice versa*; on the other hand, incomparability has no behavioural implications, it merely records the absence of a(n expressed) behavioural disposition to choose between f and g.

3. MONOIDS OF TRANSFORMATIONS

Symmetry is generally characterised mathematically as invariance under certain transformations. In this section, we provide the necessary mathematical apparatus that will allow us to describe and characterise symmetry for the belief models we are interested in.

3.1. **Transformations and lifting.**
We are interested in models for beliefs that concern a random variable X. So let us begin by concentrating on transformations of the set of possible values $\mathscr{X}$ for X. A *transformation* of $\mathscr{X}$ is defined mathematically as a map $T : \mathscr{X} \to \mathscr{X} : x \mapsto Tx$. At this point, we do not require that such a map T should be *onto* (or surjective), i.e., that $T(\mathscr{X}) := \{Tx : x \in \mathscr{X}\}$ should be equal to $\mathscr{X}$. Neither do we require that T should be *one-to-one* (or injective), meaning that $Tx = Ty$ implies $x = y$ for all x and y in $\mathscr{X}$. A transformation of $\mathscr{X}$ that is both onto and one-to-one will be called a *permutation* of $\mathscr{X}$, but we shall in the sequel also need to consider transformations of $\mathscr{X}$ that are not permutations.

Suppose we have two transformations, T and S, of $\mathscr{X}$ that are of interest. Then there is no real reason why we shouldn't also consider the combined action of T and S on $\mathscr{X}$, leading to new transformations $ST := S \circ T$ and $TS := T \circ S$, defined by $(ST)x := S(Tx)$ and similarly $TSx := T(Sx)$ for all x in $\mathscr{X}$. And of course, we could also consider in a similar way TST and STS, or for that matter $TTTSST$, which we shall also write as T^3S^2T. So it is natural in this context to consider a set $\mathscr{T}$ of transformations of $\mathscr{X}$ that is closed under composition, i.e.,

$$(\forall T, S \in \mathscr{T})(TS \in \mathscr{T}) \tag{SG}$$

Such a set is called a *semigroup of transformations*.[22] If moreover the semigroup $\mathscr{T}$ contains the identity map $\mathrm{id}_{\mathscr{X}}$, defined by $\mathrm{id}_{\mathscr{X}} x := x$ for all x in $\mathscr{X}$, it is called a *monoid*. As the identity map leaves all elements of $\mathscr{X}$ unchanged, it has no implications as far as symmetry and invariance are concerned, and we can therefore in what follows assume without loss of generality that any $\mathscr{T}$ we consider actually contains $\mathrm{id}_{\mathscr{X}}$ (is a monoid).

A monoid $\mathscr{T}$ is *Abelian* if $ST = TS$ for all T and S in $\mathscr{T}$. An important example of an Abelian monoid is the following. Consider a single transformation T of $\mathscr{X}$, and the Abelian monoid $\mathscr{T}_T$ generated by T, given by

$$\mathscr{T}_T := \{T^n \colon n \geq 0\},$$

where $T^0 := \mathrm{id}_{\mathscr{X}}$ is the identity map on $\mathscr{X}$, $T^1 := T$ and for $n \geq 2$,

$$T^n := \underbrace{T \circ T \circ \cdots \circ T}_{n \text{ times}}.$$

A monoid $\mathscr{T}$ of transformations is called *left-* (respectively *right-*)*cancellable* when for every transformation T in $\mathscr{T}$ there is some S in $\mathscr{T}$ such that $ST = \mathrm{id}_{\mathscr{X}}$ (respectively $TS = \mathrm{id}_{\mathscr{X}}$). This transformation S is then called a *left-* (respectively *right-*)*inverse* of T. If $\mathscr{T}$ is both left- and right-cancellable, then the left-and right-inverses of T are unique and coincide for any T in $\mathscr{T}$, and $\mathscr{T}$ is called a *group*. Any element of $\mathscr{T}$ is then a permutation of $\mathscr{X}$.

For our purposes here, we generally only need to assume that $\mathscr{T}$ is a monoid, because there interesting (and relevant) situations where $\mathscr{T}$ is not a group; this is for instance the case for the Abelian monoid of the *shift* transformations of the set of natural numbers $\mathbb{N}$:

$$\mathscr{T}_\theta := \{\theta^n \colon n \geq 0\}, \tag{1}$$

where $\theta(m) = m + 1$, and $\theta^n(m) = m + n$ for all natural numbers m and n. Another important example is the monoid $\mathscr{T}_{\mathscr{X}}$ of all transformations of $\mathscr{X}$, which is generally not Abelian, nor a group.

Since we are also concerned with gambles f on $\mathscr{X}$, we need a way to turn a transformation of $\mathscr{X}$ into a transformation of $\mathscr{L}(\mathscr{X})$. This is done by the procedure of *lifting*: given any gamble f on $\mathscr{X}$, we shall denote by $T^t f$ the gamble $f \circ T$, i.e.,

$$T^t f(x) := f(Tx),$$

for all x in $\mathscr{X}$. For an event A, $T^t I_A = I_{T^{-1}(A)}$, where $T^{-1}(A) := \{x \in \mathscr{X} : Tx \in A\}$ is the so-called *inverse image* of A under T. On the other hand, given a constant μ, we have $T^t \mu = \mu$ for any transformation T.

The following observation is quite important. Consider two transformations T and S on $\mathscr{X}$. Then for any gamble f on $\mathscr{X}$ we see that

$$(ST)^t f = f \circ (S \circ T) = (f \circ S) \circ T = (S^t f) \circ T = T^t (S^t f),$$

so $(ST)^t = T^t S^t$, and lifting reverses the order of application of the transformations: for x in $\mathscr{X}$, STx means that T is applied first to x, and then S to Tx. For f in $\mathscr{L}(\mathscr{X})$, $(ST)^t f$ means that S^t is applied first to f and then T^t to $S^t f$.

Any transformation T of $\mathscr{X}$ can therefore be lifted to a transformation T^t of $\mathscr{L}(\mathscr{X})$, and we denote the corresponding set of liftings by $\mathscr{T}^t$. $\mathscr{T}^t$ is then a monoid of transformations of $\mathscr{L}(\mathscr{X})$. Lifting preserves the most common properties of semigroups, taking into account the above-mentioned order-inversion: being a monoid, being Abelian, and being a group are preserved under lifting. But being left-cancellable is turned into being

---

[22]A semigroup is defined as a set with a binary operation that is internal and associative. Composition of maps is always an associative binary operation, and (SG) guarantees that it is internal in $\mathscr{T}$.

right-cancellable, and *vice versa*. Lifting also has the interesting property that it turns a transformation T on $\mathscr{X}$ into a *linear* transformation T^t of the linear space $\mathscr{L}(\mathscr{X})$: for any pair of gambles f and g on $\mathscr{X}$ and any real numbers λ and μ, we have

$$T^t(\lambda f + \mu g) = \lambda T^t f + \mu T^t g.$$

3.2. **Invariant (sets of) gambles.** We now turn to the important notions of invariance under transformations. We start with the invariance of a set of gambles, because that is the most general notion, from which all other notions of invariance can be derived. If $\mathscr{K}$ is a set of gambles on $\mathscr{X}$, and T any transformation of $\mathscr{T}$, then we denote by

$$T^t \mathscr{K} := \left\{ T^t f \colon f \in \mathscr{K} \right\}$$

the direct image of the set $\mathscr{K}$ under T^t, and we say that $\mathscr{K}$ *is $\mathscr{T}$-invariant* if

$$(\forall T^t \in \mathscr{T}^t)(T^t \mathscr{K} \subseteq \mathscr{K}),$$

i.e., if all transformations in $\mathscr{T}^t$ are *internal* in $\mathscr{K}$.[23]

A gamble f on $\mathscr{X}$ is called *$\mathscr{T}$-invariant* if the singleton $\{f\}$ is, i.e., if $T^t f = f$ for all transformations T in the monoid $\mathscr{T}$. We call an event A $\mathscr{T}$-invariant if its indicator I_A is, i.e., if $T^{-1}(A) = A$ for all T in $\mathscr{T}$.

Let us denote by $\mathscr{I}_{\mathscr{T}}$ the set of all $\mathscr{T}$-invariant events. It is easy to check that $\mathscr{I}_{\mathscr{T}}$ is an *ample field*, i.e., it contains $\emptyset$ and $\mathscr{X}$, and it is closed under arbitrary unions and complementation, and therefore also under arbitrary intersections. For any x in $\mathscr{X}$, we shall call

$$[x]_{\mathscr{T}} := \bigcap \{A \colon A \in \mathscr{I}_{\mathscr{T}} \text{ and } x \in A\}$$

the *$\mathscr{T}$-invariant atom containing x*. It is the smallest $\mathscr{T}$-invariant event that contains x. Any $\mathscr{T}$-invariant event A is a union of $\mathscr{T}$-invariant atoms: $A = \bigcup_{x \in A} [x]_{\mathscr{T}}$. We shall denote by $\mathscr{A}_{\mathscr{T}}$ the set of all invariant atoms: $\mathscr{A}_{\mathscr{T}} := \{[x]_{\mathscr{T}} \colon x \in \mathscr{X}\}$. It is a partition of $\mathscr{X}$. A gamble f on $\mathscr{X}$ is $\mathscr{T}$-invariant if and only if it is constant on the $\mathscr{T}$-invariant atoms of $\mathscr{X}$.

Of course, the bigger the set of transformations $\mathscr{T}$, the smaller the number of $\mathscr{T}$-invariant events (or, equivalently, the bigger the atoms $[x]_{\mathscr{T}}$). The following proposition relates the $\mathscr{T}$-invariant atoms $[x]_{\mathscr{T}}$ to the images of x under the transformations in $\mathscr{T}$.

Proposition 1. *Let $\mathscr{T}$ be a monoid of transformations of $\mathscr{X}$, and let x be any element of $\mathscr{X}$. In general we have that $\{Tx \colon T \in \mathscr{T}\} \subseteq [x]_{\mathscr{T}}$. If $\mathscr{T}$ is left-cancellable, then $[x]_{\mathscr{T}} = \{Tx \colon T \in \mathscr{T}\}$.*

Proof. Fix x in $\mathscr{X}$. Let $\mathscr{T}(x) := \{Tx \colon T \in \mathscr{T}\}$ for brevity of notation. Consider any T in $\mathscr{T}$. Since $[x]_{\mathscr{T}}$ is T-invariant, we have that $T^{-1}([x]_{\mathscr{T}}) = [x]_{\mathscr{T}}$. Since $x \in [x]_{\mathscr{T}}$ because $\mathscr{T}$ is a monoid, we infer from this equality that $Tx \in [x]_{\mathscr{T}}$. Hence indeed $\mathscr{T}(x) \subseteq [x]_{\mathscr{T}}$.

To prove the converse inequality, assume that $\mathscr{T}$ is left-cancellable. Consider any S in $\mathscr{T}$. If we can prove that $\mathscr{T}(x)$ is S-invariant, meaning that $S^{-1}(\mathscr{T}(x)) = \mathscr{T}(x)$, then the proof is complete, since then $\mathscr{T}(x)$ will be $\mathscr{T}$-invariant, and since this set contains x [because $\mathrm{id}_{\mathscr{X}} \in \mathscr{T}$], it must include the smallest $\mathscr{T}$-invariant set $[x]_{\mathscr{T}}$ that contains x. So we set out to prove that $S^{-1}(\mathscr{T}(x)) = \mathscr{T}(x)$. Consider any y in $\mathscr{X}$. First assume that $y \in \mathscr{T}(x)$. Then there is some T in $\mathscr{T}$ such that $y = Tx$, whence $Sy = STx \in \mathscr{T}(x)$, since $ST \in \mathscr{T}$. Conversely, assume that $y \in S^{-1}(\mathscr{T}(x))$, or equivalently, that $Sy \in \mathscr{T}(x)$, then there is some T in $\mathscr{T}$ such that $Sy = Tx$, and since $\mathscr{T}$ is assumed to be left-cancellable, there is some S' in $\mathscr{T}$ such that $S'S = \mathrm{id}_{\mathscr{X}}$, whence $\mathscr{T}(x) \ni S'Tx = S'Sy = y$, since $S'T \in \mathscr{T}$. $\qquad\square$

---

[23]So $\mathscr{T}^t$ is a monoid of transformations of $\mathscr{K}$.

An important special case is the following. Consider a transformation T of $\mathscr{X}$, and the Abelian monoid $\mathscr{T}_T = \{T^n \colon n \geq 0\}$ generated by T. Then a set of gambles $\mathscr{K}$ is $\mathscr{T}_T$-invariant if and only if $T^t \mathscr{K} \subseteq \mathscr{K}$, and we simply say that $\mathscr{K}$ is T-*invariant*. Similarly, a gamble f is $\mathscr{T}_T$-invariant if and only if $T^t f = f$, and we say that f is T-*invariant*. In what follows, we shall always use the phrase 'T-invariant' for '$\mathscr{T}_T$-invariant'. Also $\mathscr{I}_T$ is the set of T-invariant events, and it is an ample field whose atoms are denoted by $[x]_T$. With this notation, we have for an arbitrary monoid $\mathscr{T}$ that $\mathscr{I}_{\mathscr{T}} = \bigcap_{T \in \mathscr{T}} \mathscr{I}_T$.

For instance, the particular case of the shift transformations of $\mathbb{N}$ given by Eq. (1) concerns the Abelian monoid generated by θ. Here, the only θ- (or shift-)invariant events are $\emptyset$ and $\mathbb{N}$, and consequently a gamble f on $\mathbb{N}$ is θ-invariant if and only if it is constant. This also shows that the equality in the first part of Proposition 1 need not hold when the monoid of transformations $\mathscr{T}$ is not left-cancellable: in the present case, we have that $\mathscr{T}_\theta(m) = \{\theta^n(m) \colon n \geq 0\} = \{n \in \mathbb{N} \colon n \geq m\}$ is strictly included in the invariant atom $[m]_\theta = \mathbb{N}$ for all $m \geq 1$.

Another interesting case is that of $\mathscr{T}_{\mathscr{X}}$, the class of all transformations of $\mathscr{X}$. This a monoid, but it is not generally a group, nor Abelian. Moreover, it is not generally left-cancellable. We have, for any element x of $\mathscr{X}$ that $\{Tx \colon T \in \mathscr{T}_{\mathscr{X}}\} = \mathscr{X}$, and from Proposition 1 we deduce in a trivial manner that $[x]_{\mathscr{T}_{\mathscr{X}}} = \mathscr{X}$: the only invariant events under all transformations of $\mathscr{X}$ are $\emptyset$ and $\mathscr{X}$. This shows that the left-cancellability condition in the second part of Proposition 1 is not generally necessary.

4. Symmetry and invariance for belief models

We now have the necessary mathematical tools for studying the issue of symmetry in relation to the belief models discussed in Section 2. We shall see that for these coherent sets of almost-desirable gambles, there is an important distinction between the concepts 'symmetry of models' (which we shall call weak invariance) and 'models of symmetry' (which we shall call strong invariance). Let us first turn to the discussion of symmetrical belief models.

4.1. **Weak invariance: symmetry of models.** Consider a monoid $\mathscr{T}$ of transformations of $\mathscr{X}$. We want to express that a belief model about the value that the random variable X assumes in $\mathscr{X}$, exhibits a symmetry that is characterised by the transformations in $\mathscr{T}$. Thus, the notion of (weak) invariance of belief models that we are about to introduce is in a sense a purely mathematical one: it expresses that these belief models are left invariant under the transformations in $\mathscr{T}$.

Definition 1 (Weak invariance)**.** A coherent set of almost-desirable gambles $\mathscr{D}$ is called *weakly $\mathscr{T}$-invariant* if it is $\mathscr{T}$-invariant as a set of gambles, i.e., if $T^t \mathscr{D} \subseteq \mathscr{D}$ for all T in $\mathscr{T}$.

Why don't we require equality rather than the weaker requirement of set inclusion in this definition? In linear algebra, invariance of a subset of a linear space with respect to a linear transformation of that space is generally defined using only the inclusion. If we recall from Section 3 that lifting turns any transformation T of $\mathscr{X}$ into a linear transformation T^t of the linear space $\mathscr{L}(\mathscr{X})$, we see that our definition of invariance is just a special case of a notion that is quite common in the mathematical literature.

A few additional comments are in order. First of all, any coherent set of almost-desirable gambles is weakly $\mathrm{id}_{\mathscr{X}}$-invariant, so we may indeed always assume without loss of generality that $\mathscr{T}$ is at least a monoid (contains $\mathrm{id}_{\mathscr{X}}$).

Secondly, we have given an invariance definition for almost-desirability, but the definition for coherent sets of really desirable gambles $\mathscr{R}$ is completely analogous: for all T in $\mathscr{T}$, $T^t\mathscr{R} \subseteq \mathscr{R}$. Observe that if $\mathscr{R}$ is weakly $\mathscr{T}$-invariant then the associated set of almost-desirable gambles $\mathscr{D}_{\mathscr{R}}$, given by (D-M), is weakly $\mathscr{T}$-invariant as well.

Thirdly, if $\mathscr{T}$ is a group (or at least left-cancellable), then the weak invariance condition is actually equivalent to $T^t\mathscr{D} = \mathscr{D}$ for all T in $\mathscr{T}$: given a transformation T in $\mathscr{T}$ and its (left-)inverse $S \in \mathscr{T}$, consider $f \in \mathscr{D}$; then $T^t(S^tf) = (ST)^tf = f$, so there is a gamble $g = S^tf$, which belongs to $\mathscr{D}$ by weak invariance, such that $f = T^tg$; this means that $f \in T^t\mathscr{D}$, so $\mathscr{D} \subseteq T^t\mathscr{D}$ as well.

In summary, weak invariance is a mathematical notion that states that a subject's behavioural dispositions, as represented by a belief model $\mathscr{D}$, are invariant under certain transformations. If we posit that a subject's dispositions are in some way a reflection of the evidence available to her, we see that weak invariance is a way to model 'symmetry of evidence'. The following examples try to argue that if there is 'symmetry of evidence', then corresponding belief models should at least be weakly invariant.

The example of shift transformations. Suppose our subject is completely ignorant about the value of a random variable X that assumes only non-negative integer values, so $\mathscr{X} = \mathbb{N}$. If her belief model is to be a reflection of the available evidence (none), we should like it to be weakly invariant with respect to the shift transformations in $\mathscr{T}_\theta$ (which is an Abelian monoid, but not a group). Indeed, if she is ignorant about X, she is also ignorant about $\theta(X) = X + 1$, apart from the fact that she knows that $\theta(X)$ cannot assume the value 0, whereas X can. Therefore, if our subject almost-accepts a gamble f, she should almost-accept θ^tf: $\theta^tf(X) = f(X+1)$ may assume the same values as $f(X)$, apart from the value $f(0)$, and because of her ignorance, our subject has no reason to treat the shifted gamble differently. ♦

The dice example. Let us go back to the die example. Suppose that whatever evidence our subject has about the outcome X of rolling the die, is left invariant by permutations π of $\mathscr{X}_6 = \{1,\ldots,6\}$. Assume that our subject almost-accepts a gamble f, meaning that she is willing to accept the uncertain reward $f(X) + \varepsilon$ for any $\varepsilon > 0$. But since the evidence gives our subject no reason to distinguish between the random variables X and $\pi(X)$, she should also be willing to accept the uncertain reward $f(\pi(X)) + \varepsilon$ for any $\varepsilon > 0$, or in other words, she should almost-accept the gamble π^tf.

We now investigate the corresponding notions for weak invariance for the equivalent belief models: coherent lower previsions and weak*-closed convex sets of coherent previsions. In order to do this, it is convenient to define the transformation of a (lower) prevision under a transformation T on $\mathscr{X}$, by lifting T to yet a higher level.

Definition 2 (Transformation of a functional). [24] Let T be a transformation of $\mathscr{X}$ and let Λ be a real-valued functional defined on a T-invariant set of gambles $\mathscr{K} \subseteq \mathscr{L}(\mathscr{X})$. Then the transformation $T\Lambda$ of Λ is the real-valued functional defined on $\mathscr{K}$ by $T\Lambda := \Lambda \circ T^t$, or equivalently, by $T\Lambda(f) := \Lambda(T^tf) = \Lambda(f \circ T)$ for all gambles f in $\mathscr{K}$.

Theorem 2. *Let $\underline{P}$ be a coherent lower prevision on $\mathscr{L}(\mathscr{X})$, $\mathscr{D}$ a coherent set of almost-desirable gambles, and $\mathscr{M}$ a weak*-closed convex set of coherent previsions on $\mathscr{L}(\mathscr{X})$. Assume that these belief models are equivalent, in the sense that they correspond to one*

----

[24]We use the same notation T for the transformation of $\mathscr{X}$ and for the corresponding transformation of a functional, first of all because we do not want to overload the mathematical notation, and also because, in contrast with lifting only once, lifting twice preserves the order of application of transformations.

another using the bijective relations in Table 1. Then the following statements are equivalent.

1. $\mathscr{D}$ *is weakly $\mathscr{T}$-invariant, in the sense that $T^t\mathscr{D} \subseteq \mathscr{D}$ for all T in $\mathscr{T}$.*
2. $\underline{P}$ *is* weakly $\mathscr{T}$-invariant, *in the sense that $T\underline{P} \geq \underline{P}$ for all T in $\mathscr{T}$, or equivalently $\underline{P}(T^t f) \geq \underline{P}(f)$ for all T in $\mathscr{T}$ and f in $\mathscr{L}(\mathscr{X})$;*
3. $\mathscr{M}$ *is* weakly $\mathscr{T}$-invariant, *in the sense that $T\mathscr{M} \subseteq \mathscr{M}$ for all T in $\mathscr{T}$, or equivalently, $TP \in \mathscr{M}$ for all P in $\mathscr{M}$ and all T in $\mathscr{T}$.*[25]

Proof. We give a circular proof. Assume that $\mathscr{D}$ is weakly $\mathscr{T}$-invariant. Consider any T in $\mathscr{T}$ and f in $\mathscr{K}$, and observe that for the corresponding lower prevision $\underline{P}$

$$\underline{P}(T^t f) = \max\left\{\mu : T^t f - \mu \in \mathscr{D}\right\} \geq \max\left\{\mu : f - \mu \in \mathscr{D}\right\} = \underline{P}(f),$$

where the inequality follows from the invariance assumption on $\mathscr{D}$. This shows that the first statement implies the second.

Next, assume that $\underline{P}$ is weakly $\mathscr{T}$-invariant, and consider any T in $\mathscr{T}$ and P in the corresponding $\mathscr{M} = \mathscr{M}(\underline{P}) = \{P : (\forall f)P(f) \geq \underline{P}(f)\}$. Then for any gamble f on $\mathscr{X}$ we have that $TP(f) = P(T^t f) \geq \underline{P}(T^t f) \geq \underline{P}(f)$, where the second inequality follows for the invariance assumption on $\underline{P}$. This tells us that indeed $TP \in \mathscr{M}(\underline{P})$, so the second statement implies the third.

Finally, assume that $\mathscr{M}$ is weakly $\mathscr{T}$-invariant. Consider any T in $\mathscr{T}$ and any gamble f in the corresponding $\mathscr{D} = \mathscr{D}_{\mathscr{M}} = \{f : (\forall P \in \mathscr{M})P(f) \geq 0\}$. Then we have for any P in $\mathscr{M}$ that $P(T^t f) = TP(f) \geq 0$, since TP belongs to $\mathscr{M}(\underline{P})$ by the invariance assumption on $\mathscr{M}$. Consequently $T^t f \in \mathscr{D}$, which proves that the third statement implies the first. $\square$

A coherent prevision P on $\mathscr{L}(\mathscr{X})$ is weakly $\mathscr{T}$-invariant if and only if $TP = P$ for all T in $\mathscr{T}$. This is easiest to prove by observing that $\mathscr{M}(P) = \{P\}$.[26] So for coherent previsions, we have an equality in the weak invariance condition. As we argued before, we generally won't have such an equality for arbitrary monoids $\mathscr{T}$, but the following corollary gives another sufficient condition on $\mathscr{T}$.

Corollary 3. *If the monoid $\mathscr{T}$ is left-cancellable, then the first weak invariance condition in Theorem 2 becomes $T^t\mathscr{D} = \mathscr{D}$ for all T in $\mathscr{T}$. If $\mathscr{T}$ is right-cancellable, then the second and third weak invariance conditions become $T\underline{P} = \underline{P}$ and $T\mathscr{M} = \mathscr{M}$ for all T in $\mathscr{T}$.*[27]

Proof. We have already proven the first statement near the beginning of Section 4.1. To prove the second statement, it suffices to show that when $\mathscr{T}$ is right-cancellable, $\mathscr{T}$-invariance implies that $\underline{P} \geq T\underline{P}$ and $\mathscr{M} \subseteq T\mathscr{M}$ for all T in $\mathscr{T}$. Consider any transformation T in the monoid $\mathscr{T}$, and let R be a right-inverse for T, i.e., $TR = \mathrm{id}_{\mathscr{X}}$. Consider a gamble h on $\mathscr{X}$, then $\underline{P}(h) = \underline{P}((TR)^t h) = \underline{P}(R^t(T^t h)) \geq \underline{P}(T^t h)$, where the inequality follows from the weak invariance of $\underline{P}$. So indeed, $\underline{P} \geq T\underline{P}$. Similarly, consider P in $\mathscr{M}$. Then $RP \in \mathscr{M}$ by weak invariance, and for any gamble f on $\mathscr{X}$, $T(RP)(f) = RP(T^t f) = P(R^t(T^t f)) =$

---

[25]This shows that our notion of a weakly invariant belief model corresponds to Pericchi and Walley's [1991] notion of a 'reasonable (or invariant) class of priors', rather than a 'class of reasonable (or invariant) priors', the latter being what our notion of strong invariance will correspond to. On the other hand, Walley [1991, Definition 3.5.1] defines a $\mathscr{T}$-invariant lower prevision $\underline{P}$ as one for which $\underline{P}(T^t f) = \underline{P}(f)$ for all $T \in \mathscr{T}$ and all gambles f, so he requires equality rather than inequality, as we do here.

[26]See Proposition 7 for a more direct proof.

[27]The reason for the difference in terms of left- versus right-cancellability lies of course in the fact that in the first condition, we work with transformations T^t of gambles, and in the second and third condition we work with transformations T of functionals, which are liftings of the former; simply recall that lifting reverses the order of application of transformations.

$P(f)$ since $R^t(T^t f) = (TR)^t f = f$. So there is a $Q = RP$ in $\mathcal{M}$ such that $P = TQ$, meaning that $P \in T\mathcal{M}$. So indeed $\mathcal{M} \subseteq T\mathcal{M}$. $\qquad\square$

We see from the definition that if a coherent set of almost-desirable gambles $\mathscr{D}$ (or a coherent lower prevision, or a weak*-closed convex set of coherent previsions) is weakly $\mathscr{T}$-invariant, it is also weakly $\mathscr{T}'$-invariant for any sub-monoid of transformations $\mathscr{T}' \subseteq \mathscr{T}$. Hence, as we add transformations, the collection of weakly invariant belief models will not increase. The limit case is when we consider the class $\mathscr{T}_{\mathscr{X}}$ of all transformations on $\mathscr{X}$. The following theorem shows that the vacuous belief models are the only ones that are *completely weakly invariant*, i.e., weakly $\mathscr{T}_{\mathscr{X}}$-invariant.

Theorem 4. *Let $\mathscr{T}_{\mathscr{X}}$ be the monoid of all transformations of $\mathscr{X}$. Then the vacuous coherent set of almost-desirable gambles $\mathscr{C}_+$ (or equivalently, the vacuous lower prevision $\underline{P}_{\mathscr{X}}$, or equivalently, the weak*-closed convex set of all coherent previsions $\mathbb{P}(\mathscr{X})$) is the only coherent set of almost-desirable gambles (coherent lower prevision, weak*-closed convex set of coherent previsions) that is weakly $\mathscr{T}_{\mathscr{X}}$-invariant.*

Proof. We give the proof for coherent sets of almost-desirable gambles. It is obvious that $\mathscr{C}_+$ is $\mathscr{T}_{\mathscr{X}}$-invariant. So, consider any $\mathscr{T}_{\mathscr{X}}$-invariant coherent set of almost-desirable gambles $\mathscr{D}$. It follows from coherence [axiom (M2)] that $\mathscr{C}_+ \subseteq \mathscr{D}$. Assume *ex absurdo* that $\mathscr{C}_+ \subset \mathscr{D}$ and let f be any gamble in $\mathscr{D} \setminus \mathscr{C}_+$. This means that there is some x_0 in $\mathscr{X}$ such that $f(x_0) < 0$. Consider the transformation T_{x_0} of $\mathscr{X}$ that maps all elements of $\mathscr{X}$ to x_0, then $T_{x_0}^t f = f(x_0)$ and it follows from the T_{x_0}-invariance of $\mathscr{D}$ that the constant gamble $f(x_0) \in \mathscr{D}$, which violates coherence axiom (M1), so $\mathscr{D}$ cannot be coherent, a contradiction.[28] $\qquad\square$

This result also tells us in particular that the vacuous belief model is always $\mathscr{T}$-invariant for any monoid of transformations $\mathscr{T}$. This implies that for any monoid of transformations $\mathscr{T}$, there always are $\mathscr{T}$-invariant belief models.

What are the behavioural consequences of weak invariance with respect to a monoid of transformations $\mathscr{T}$? It seems easiest to study this in terms of coherent lower previsions. First of all, we have that for any gamble f on $\mathscr{X}$ and any T in $\mathscr{T}$, our subject's supremum buying price $\underline{P}(T^t f)$ for the transformed gamble $T^t f$ should not be strictly smaller that her supremum price $\underline{P}(f)$ for buying f itself.

But there is also a more interesting consequence. Indeed, it follows from the coherence of $\underline{P}$ that

$$\underline{P}(f - T^t f) \le \underline{P}(f) - \underline{P}(T^t f) \le 0.$$

Walley [1991, Section 3.8.1] suggests that a subject *strictly prefers* a gamble f to a gamble g, which we denote as $f \succ g$, if $f > g$, or also if she accepts to pay some (strictly) positive price for exchanging g with f, so if $\underline{P}(f - g) > 0$. This means that weak $\mathscr{T}$-invariance implies that

$$f \not\succ T^t f \text{ for all } f \text{ in } \mathscr{L}(\mathscr{X}) \text{ and all } T \text{ in } \mathscr{T} \text{ such that } f \not> T^t f$$

which models that our subject *has no reason* (or disposition) *to strictly prefer any gamble f to any of its transformations $T^t f$ that it doesn't strictly dominate.*

---

[28]A similar argument tells us that the same result holds for complete weak invariance of coherent sets of really desirable gambles, where now the axiom (D1) will be violated.

4.2. **Strong invariance: models of symmetry.** Next, suppose that our subject believes that the (phenomenon underlying the) random variable X is subject to symmetry with respect to the transformations T in $\mathscr{T}$, so that she has *reason not to distinguish* between a gamble f and its transformation $T^t f$. Let us give an example to get a more intuitive understanding of what this means.

The dice example (cont.). Again, let us go back to the die example. Consider the gambles $I_{\{x\}}$, for $x \in \mathscr{X}_6 := \{1,\ldots,6\}$. Since our subject believes the die (and the rolling mechanism behind it) to be symmetrical, she will be willing to exchange any gamble $I_{\{x\}}$ for any other gamble $I_{\{y\}}$ in return for any strictly positive amount of utility: $I_{\{x\}} - I_{\{y\}}$ should therefore be almost-desirable to her, or in other words, in terms of her lower prevision $\underline{P}$:

$$\underline{P}(I_{\{x\}} - I_{\{y\}}) \geq 0 \text{ for all } x \text{ and } y \text{ in } \mathscr{X}_6.$$

This is equivalent to stating that $I_{\{x\}} - \pi^t I_{\{x\}}$ should be almost-desirable, or that $\underline{P}(I_{\{x\}} - \pi^t I_{\{x\}}) \geq 0$ for all $x \in \mathscr{X}_6$ and all permutations π of $\mathscr{X}_6$. Now the only coherent lower prevision that satisfies these requirements is the uniform (precise) prevision, which assigns precise probability $\frac{1}{6}$ to each event $\{x\}$ [simply observe that for any coherent prevision P in $\mathscr{M}(\underline{P})$ it follows from these requirements that $P(I_{\{x\}}) = P(I_{\{y\}})$]. ◆

Let us now try and formalise the intuitive requirements in this example into a more formal definition. We stated above that if our subject believes that the (phenomenon underlying the) random variable X is subject to symmetry with respect to the transformations T in $\mathscr{T}$, then she has *reason not to distinguish* between a gamble f and its transformation $T^t f$. Suppose she has the gamble f in her possession, then she should be willing to exchange this for the gamble $T^t f$ in return for any strictly positive price, and *vice versa*. This means that she should almost-accept both $f - T^t f$ and $T^t f - f$, or in the language of Section 2.5, that she is *indifferent between f and $T^t f$*: $f \approx T^t f$. If $\mathscr{D}$ is her coherent set of almost-desirable gambles, this means that

$$f - T^t f \in \mathscr{D} \text{ and } T^t f - f \in \mathscr{D} \text{ for all } f \text{ in } \mathscr{L}(\mathscr{X}) \text{ and all } T \text{ in } \mathscr{T}.$$

If we define

$$\mathscr{D}_\mathscr{T} := \left\{ f - T^t f \colon f \in \mathscr{L}(\mathscr{X}), T \in \mathscr{T} \right\} = \left\{ T^t f - f \colon f \in \mathscr{L}(\mathscr{X}), T \in \mathscr{T} \right\},$$

this leads to the following definition.

Definition 3. A coherent set of almost-desirable gambles $\mathscr{D}$ is called *strongly $\mathscr{T}$-invariant* if $f - T^t f \in \mathscr{D}$ for all f in $\mathscr{L}(\mathscr{X})$ and all T in $\mathscr{T}$, or equivalently, if $\mathscr{D}_\mathscr{T} \subseteq \mathscr{D}$.

The following theorem gives equivalent characterisations of strong invariance in terms of the alternative types of belief models.

Theorem 5. *Let $\underline{P}$ be a coherent lower prevision on $\mathscr{L}(\mathscr{X})$, $\mathscr{D}$ a coherent set of almost-desirable gambles, and $\mathscr{M}$ a weak*-closed convex set of coherent previsions on $\mathscr{L}(\mathscr{X})$. Assume that these belief models are equivalent, in the sense that they correspond to one another using the bijective relations in Table 1. Then the following statements are equivalent:*

1. *$\mathscr{D}$ is strongly $\mathscr{T}$-invariant, in the sense that $\mathscr{D}_\mathscr{T} \subseteq \mathscr{D}$;*
2. *$\underline{P}$ is strongly $\mathscr{T}$-invariant, in the sense that $\underline{P}(f - T^t f) \geq 0$ and $\underline{P}(T^t f - f) \geq 0$, and therefore $\underline{P}(f - T^t f) = \underline{P}(T^t f - f) = 0$ for all f in $\mathscr{L}(\mathscr{X})$ and T in $\mathscr{T}$;*
3. *$\mathscr{M}$ is strongly $\mathscr{T}$-invariant, in the sense that $TP = P$ for all P in $\mathscr{M}$ and all T in $\mathscr{T}$.*[29]

---

[29]So strongly invariant belief models correspond to the Pericchi and Walley's [1991] notion of a 'class of reasonable (or invariant) priors'.

Proof. We give a circular proof. Assume that $\mathscr{D}$ is strongly $\mathscr{T}$-invariant, and consider any gamble f on $\mathscr{X}$ and any T in $\mathscr{T}$. Then we find for the associated coherent lower prevision $\underline{P}$ that $\underline{P}(f - T^t f) = \max\{s\colon f - T^t f - s \in \mathscr{D}\} \geq 0$, and similarly that $\underline{P}(T^t f - f) \geq 0$. But since $\underline{P}$ is coherent, we find that also $\underline{P}(f - T^t f) = -\overline{P}(T^t f - f) \leq -\underline{P}(T^t f - f) \leq 0$ and similarly $\underline{P}(T^t f - f) = -\overline{P}(f - T^t f) \leq -\underline{P}(f - T^t f) \leq 0$, whence indeed $\underline{P}(f - T^t f) = \underline{P}(T^t f - f) = 0$, so the first statement implies the second.

Next, assume that $\underline{P}$ is strongly $\mathscr{T}$-invariant and consider any P in the associated set of dominating coherent previsions $\mathscr{M} = \{P\colon (\forall f)(P(f) \geq \underline{P}(f))\}$ and any T in $\mathscr{T}$. Then for any gamble f on $\mathscr{X}$ we see that $P(f - T^t f) \geq 0$ and $P(T^t f - f) \geq 0$, and since P is a coherent prevision, this implies that $P(T^t f) = P(f)$, so indeed $TP = P$. Hence, the second statement implies the third.

Finally, assume that $\mathscr{M}$ is strongly $\mathscr{T}$-invariant, and consider any gamble f on $\mathscr{X}$ and any T in $\mathscr{T}$. Then for all P in $\mathscr{M}$ we have that $P(f - T^t f) = P(T^t f - f) = 0$, so both $f - T^t f$ and $T^t f - f$ belong to the associated set of almost-desirable gambles $\mathscr{D} = \{g\colon (\forall P \in \mathscr{M})P(g) \geq 0\}$. This tells us that the third statement implies the first. $\qquad\square$

Let us now study in more detail the relationship between weak and strong invariance. First of all, strong invariance implies weak invariance, but generally not the other way around. It is easiest to see this using weak*-closed convex sets of coherent previsions $\mathscr{M}$. If $\mathscr{M}$ is strongly $\mathscr{T}$-invariant, we have that $TP = P$ and consequently $TP \in \mathscr{M}$ for all P in $\mathscr{M}$, so $\mathscr{M}$ is also weakly $\mathscr{T}$-invariant. To see that the converse doesn't generally hold, consider the set of all coherent previsions $\mathbb{P}(\mathscr{X})$ (the vacuous belief model), which is weakly invariant with respect to any monoid of transformations, but not necessarily strongly so, as, unless $\mathscr{X}$ contains only one element, we can easily find transformations T and coherent previsions P such that TP is different from P (also see Theorem 6 below).

But the theorem above, when interpreted well, also tells us a number of very interesting things on this issue. First of all, we see that a coherent prevision P on $\mathscr{L}(\mathscr{X})$ is strongly $\mathscr{T}$-invariant if and only if it is weakly $\mathscr{T}$-invariant, so both notions of invariance coincide for coherent previsions. *So anyone who insists on modelling beliefs with Bayesian belief models (coherent previsions) only, cannot distinguish between the two types of invariance.* This confirms in general what we claimed in the Introduction about Bayesian belief models. From now on, we shall therefore no longer distinguish between strong and weak invariance for coherent previsions, and simply call them *invariant*.

Furthermore, we see that a coherent lower prevision $\underline{P}$ is strongly $\mathscr{T}$-invariant if and only if all its dominating coherent lower previsions are, or equivalently, if all its dominating coherent previsions, i.e., all the coherent previsions in $\mathscr{M}(\underline{P})$, are $\mathscr{T}$-invariant. Or even stronger, it is easy to see that a coherent lower prevision is strongly invariant if and only if it is a lower envelope of some (not necessarily weak*-closed nor convex) set of invariant coherent previsions.

The notions of weak and strong invariance, and the motivation for introducing them, are tailored to the direct behavioural interpretation of lower previsions, or the equivalent belief models. But what happens if we give a lower prevision $\underline{P}$ a Bayesian sensitivity analysis interpretation? We then hold that there is some actual precise coherent prevision P_a modelling the subject's uncertainty about the random variable X, that we have only imperfect information about in the sense that we only know that $P_a \geq \underline{P}$, or equivalently, that $P_a \in \mathscr{M}(\underline{P})$. Assume that we want the imperfect model $\underline{P}$ to capture that there is 'symmetry of evidence' with respect to a monoid of transformations $\mathscr{T}$. The actual model P_a then should be weakly $\mathscr{T}$-invariant, but since this is a (precise) coherent prevision, we can

not distinguish between weak and strong invariance, and it should therefore simply be $\mathscr{T}$-invariant: $TP_a = P_a$ for all $T \in \mathscr{T}$. Since $\mathscr{M}(\underline{P})$ is interpreted as the set of candidate models for P_a, all of the coherent previsions P in $\mathscr{M}(\underline{P})$ must be $\mathscr{T}$-invariant too, or equivalently $\underline{P}$ must be *strongly* $\mathscr{T}$-invariant. A completely analogous course of reasoning shows that if we want $\underline{P}$ to capture 'evidence of symmetry', $\underline{P}$ must be strongly $\mathscr{T}$-invariant as well. So in contradistinction with the direct behavioural interpretation, *on a Bayesian sensitivity analysis interpretation of $\underline{P}$, we cannot distinguish between 'symmetry of evidence' and 'evidence of symmetry', and strong invariance is the proper symmetry property to use in both cases.*[30]

As is the case for weak invariance, a belief model that is strongly $\mathscr{T}$-invariant, is also strongly $\mathscr{T}'$-invariant for any sub-monoid $\mathscr{T}' \subseteq \mathscr{T}$. But in contrast with weak invariance, given any monoid $\mathscr{T}$, there do not always exist coherent belief models that are strongly invariant with respect to $\mathscr{T}$. This is an immediate consequence of the following theorem, which makes an even stronger claim: it is totally *irrational* to require *complete* strong invariance, i.e., strong invariance with respect to the monoid $\mathscr{T}_{\mathscr{X}}$ of all transformations of $\mathscr{X}$.

Theorem 6. *Assume that $\mathscr{X}$ contains more than one element. Then any belief model that is strongly $\mathscr{T}_{\mathscr{X}}$-invariant incurs a sure loss.*

Proof. We shall give a proof for lower previsions. Assume *ex absurdo* that $\underline{P}$ avoids sure loss, so $\mathscr{M}(\underline{P})$ is non-empty. Consider any P in $\mathscr{M}(\underline{P})$ and any non-constant gamble f on $\mathscr{X}$ [there is at least one such gamble because $\mathscr{X}$ contains more than one element]. This implies that there are (different) x_1 and x_2 in $\mathscr{X}$ such that $f(x_1) \neq f(x_2)$. For any y in $\mathscr{X}$, consider the transformation T_y that maps all elements of $\mathscr{X}$ to y. Then we find that $T_y^t f = f(y)$, whence $P(f(y) - f) \geq \underline{P}(f(y) - f) \geq 0$ and $P(f - f(y)) \geq \underline{P}(f - f(y)) \geq 0$, since $\underline{P}$ is by assumption in particular strongly T_y-invariant. Consequently $P(f) = f(y)$. But this holds in particular for $y = x_1$ and for $y = x_2$, so we infer that $f(x_1) = P(f) = f(x_2)$, a contradiction. $\qquad\square$

In fact, we easily see in this proof that given the transformation T_y that maps all elements of $\mathscr{X}$ to y, the only strongly T_y-invariant belief model that avoids sure loss is the constant prevision on y. Consequently, if we consider a monoid $\mathscr{T}$ that includes two different constant transformations, any belief model that is strongly $\mathscr{T}$-invariant incurs a sure loss.

As a result, we see that there are monoids $\mathscr{T}$ for which there are no strongly invariant coherent (lower) previsions. Under which conditions, then, are there strongly $\mathscr{T}$-invariant coherent (lower) previsions? It seems easiest, and yields most insight, if we look at this problem in terms of sets of almost-desirable gambles: indeed if we consider a coherent lower prevision $\underline{P}$ on $\mathscr{L}(\mathscr{X})$, then it is strongly $\mathscr{T}$-invariant if and only if for its associated set of almost-desirable gambles $\mathscr{D}_{\underline{P}} = \{f \in \mathscr{L}(\mathscr{X}) : \underline{P}(f) \geq 0\}$ we have that $\mathscr{D}_{\mathscr{T}} \subseteq \mathscr{D}_{\underline{P}}$. We can consider $\mathscr{D}_{\mathscr{T}}$ itself as a set of almost-desirable gambles, but at this point, we do not know whether $\mathscr{D}_{\mathscr{T}}$ is coherent, or whether it even avoids sure loss. Interestingly, the set of coherent previsions that is associated with $\mathscr{D}_{\mathscr{T}}$ is given by

$$\mathscr{M}(\mathscr{D}_{\mathscr{T}}) = \{P \in \mathbb{P}(\mathscr{X}) : (\forall g \in \mathscr{D}_{\mathscr{T}})(P(g) \geq 0)\}$$
$$= \{P \in \mathbb{P}(\mathscr{X}) : (\forall f \in \mathscr{L}(\mathscr{X}))(\forall T \in \mathscr{T})(P(f) = P(T^t f))\}.$$

So $\mathscr{M}(\mathscr{D}_{\mathscr{T}})$ is precisely the convex and weak*-closed set of all $\mathscr{T}$-invariant coherent previsions, and $\underline{P}$ is strongly $\mathscr{T}$-invariant if and only if $\mathscr{M}(\underline{P}) \subseteq \mathscr{M}(\mathscr{D}_{\mathscr{T}})$, or in other

---

[30]See [Walley, 1991, Section 9.5] for related comments about the difference between permutability and exchangeability. These notions will be briefly discussed in Section 9.2.

words, if and only if all coherent previsions that dominate $\underline{P}$ are $\mathscr{T}$-invariant. So there are strongly $\mathscr{T}$-invariant coherent lower previsions if and only if $\mathscr{M}(\mathscr{D}_{\mathscr{T}}) \neq \emptyset$, i.e., if there are $\mathscr{T}$-invariant coherent previsions, and in this case the lower envelope of $\mathscr{M}(\mathscr{D}_{\mathscr{T}})$ is the point-wise smallest strongly $\mathscr{T}$-invariant coherent lower prevision.

In summary, we see that there are $\mathscr{T}$-invariant coherent previsions if and only if the set of almost-desirable gambles $\mathscr{D}_{\mathscr{T}}$ avoids sure loss,[31] which, taking into account (M-ASL), is equivalent[32] to the condition[33]

$$\sup \sum_{k=1}^{n} \left[f_k - T_k^t f_k \right] \geq 0 \quad \text{for all } n \geq 0, f_1, \ldots, f_n \text{ in } \mathscr{L}(\mathscr{X}) \text{ and } T_1, \ldots, T_n \text{ in } \mathscr{T}. \quad (2)$$

In that case, the natural extension $\mathscr{E}_{\mathscr{T}} := \mathscr{E}_{\mathscr{D}_{\mathscr{T}}}^{m}$ of $\mathscr{D}_{\mathscr{T}}$ to a coherent set of almost-desirable gambles is given by[34]

$$\mathscr{E}_{\mathscr{T}} = \bigcap_{\varepsilon > 0} \left\{ f \in \mathscr{L}(\mathscr{X}) \colon f - \varepsilon \geq \sum_{k=1}^{n} \left[f_k - T_k^t f_k \right] \text{ for some } n \geq 0, f_k \in \mathscr{L}(\mathscr{X}), T_k \in \mathscr{T} \right\} \quad (3)$$

This is the smallest coherent and strongly $\mathscr{T}$-invariant set of almost-desirable gambles, or in other words, the belief model that represents evidence of symmetry involving the monoid $\mathscr{T}$. The corresponding lower prevision, defined by[35]

$$\underline{E}_{\mathscr{T}}(f) = \min \{ P(f) \colon P \in \mathscr{M}(\mathscr{D}_{\mathscr{T}}) \} \quad (4)$$

$$= \max \{ \mu \in \mathbb{R} \colon f - \mu \in \mathscr{E}_{\mathscr{T}} \} \quad (5)$$

is then, by virtue of Eq. (4) [see also Theorem 10 further on], the point-wise smallest (most conservative) strongly $\mathscr{T}$-invariant coherent lower prevision on $\mathscr{L}(\mathscr{X})$, and if we combine Eqs. (3) and (5), we find that[36]

$$\underline{E}_{\mathscr{T}}(f) = \sup \left\{ \inf \left[f - \sum_{k=1}^{n} \left[f_k - T_k^t f_k \right] \right] \colon n \geq 0, f_k \in \mathscr{L}(\mathscr{X}), T_k \in \mathscr{T} \right\}. \quad (6)$$

Remember that this lower prevision is only well-defined (assumes finite real values) whenever the condition (2) is satisfied. Taking into account Theorem 10 further on, we deduce that a coherent (lower) prevision is (strongly) $\mathscr{T}$-invariant if and only if it dominates $\underline{E}_{\mathscr{T}}$. Also, $\underline{E}_{\mathscr{T}}$ is the belief model we should use if nothing else but the evidence of symmetry is given. Finally, this formula for the *lower* prevision is constructive, but usually the existence of invariant previsions (on infinite spaces) is proven in a non-constructive (Hahn–Banach) way; see Section 8, and also Agnew and Morse [1938] and Bhaskara Rao and Bhaskara Rao [1983, Section 2.1.3(8)]. So we cannot usually get to the coherent invariant previsions by construction, but we can always construct their lower envelope explicitly!

---

[31] Also see Walley's [1991, Lemma 3.3.2] Separation Lemma.

[32] Observe that the set $\mathscr{D}_{\mathscr{T}}$ is a *cone*, i.e., closed under scalar multiplication with non-negative real numbers.

[33] The same condition was derived by Walley [1991, Theorem 3.5.2 and Corollary 3.5.4] using an argument that works directly with coherent lower previsions. Although our argument strongly plays on the connection between the three equivalent types of belief models of Table 1, we believe that it produces more insight, once this connection is fully understood.

[34] Again, observe that $\mathscr{D}_{\mathscr{T}}$ is a cone.

[35] It is easy to see that $\mathscr{M}(\mathscr{D}_{\mathscr{T}}) = \mathscr{M}(\mathscr{E}_{\mathscr{T}})$.

[36] Again, Walley [1991, Theorem 3.5.2 and Corollary 3.5.4] proves the same result in a different manner, see also footnote 33.

We shall have much more to say about the existence of strongly invariant belief models in Section 7, where we show that this existence is guaranteed in particular if the monoid $\mathscr{T}$ is Abelian, or if it is a finite group. The following counterexample tells us that there is no such guarantee for infinite groups.

Example 1 (Permutation invariance on the natural numbers). Consider the set $\mathscr{P}_{\mathbb{N}}$ of all permutations of the set of natural numbers $\mathbb{N}$. We show that there are no (strongly) $\mathscr{P}_{\mathbb{N}}$-invariant coherent (lower) previsions on $\mathscr{L}(\mathbb{N})$ by showing that the condition (2) doesn't hold. Indeed, consider the partition of $\mathbb{N}$ made up of the sets

$$R_3^r = \{3n + r : n \in \mathbb{N}\}, \quad r = 0, 1, 2,$$

and any permutations π_r for $r = 0, 1, 2$ such that for all $n \in \mathbb{N}$, $\pi_r(n) \in R_3^r$ if and only if $n \notin R_3^r$ [for instance, let π_r be involutive and such that it assigns the first element of R_3^r to the first of $(R_3^r)^c$, the second element of R_3^r to the second of $(R_3^r)^c$, etc.] Consider the gamble $G = \sum_{r=0}^{2}[I_{R_3^r} - \pi_r^t I_{R_3^r}]$ on $\mathbb{N}$, then we are done if we can show that $\sup G < 0$. Indeed, if $n \in R_3^r$ then $G(n) = 1 + 0 + 0 - (1 + 1 + 0) = -1$ for $r = 0, 1, 2$, so $\sup G = -1$. ♦

These results expose another fundamental difference between weak and strong invariance: while strong invariance with respect to a greater number of transformations means that we must refine our beliefs (i.e, it make them more precise), this is not the case with weak invariance.

On the other hand, strong invariance is preserved by dominating lower previsions: if $\underline{P}_1$ is a coherent lower prevision that is strongly $\mathscr{T}$-invariant and $\underline{P}_2$ is a coherent lower prevision that dominates $\underline{P}_1$, then $\underline{P}_2$ is also strongly $\mathscr{T}$-invariant. It indeed seems reasonable that, if a subject has evidence of symmetry, and she has some additional information that allows her to make her judgements more precise, she can add assessments while still preserving strong invariance. But a similar result does not hold for weak invariance: since the vacuous lower prevision is weakly $\mathscr{T}_{\mathscr{X}}$-invariant, this would mean that any lower prevision should be weakly $\mathscr{T}_{\mathscr{X}}$-invariant, *quod non*.

In summary, there is an important conceptual difference between weak and strong invariance. Weakly invariant belief models capture in particular that a subject *has no reason to* strictly prefer a gamble f to its transformation $T^t f$ whenever $f \not> T^t f$. Strong invariance captures that a subject *has reason not to* distinguish between, i.e., to be indifferent between, the gambles f and $T^t f$. And it is only if you insist on using Bayesian belief models always that you must infer indifference from having no reason to (strictly) prefer. This is of particular relevance for belief models that try to represent a subject's complete ignorance, as we now proceed to show.

5. Modelling complete ignorance

Suppose our subject is completely ignorant about the value that X assumes in $\mathscr{X}$. Then she has no relevant information that would allow her to favour one possible value of X over another. This implies that the corresponding belief model should be symmetric in the possible values of X, or in other words it should be weakly invariant with respect to the group $\mathscr{P}_{\mathscr{X}}$ of all permutations of $\mathscr{X}$. This leads to a form of Walley's [1991, Section 5.5.1] Symmetry Principle.

Symmetry Principle (SP). *If a subject is completely ignorant about the value of a random variable X in $\mathscr{X}$, then her corresponding belief model should be weakly invariant with respect to the group $\mathscr{P}_{\mathscr{X}}$ of all permutations of $\mathscr{X}$.*

We have mentioned before that the appropriate belief model for complete ignorance about X seems to be the vacuous lower prevision $\underline{P}_{\mathscr{X}}$. But SP by itself is not sufficient to single out this lower prevision: if, for instance, $\mathscr{X}$ is finite, then the uniform precise prevision P^u, given by

$$P^u(f) = \frac{1}{|\mathscr{X}|} \sum_{x \in \mathscr{X}} f(x)$$

for each gamble f on $\mathscr{X}$, which assigns equal probability mass $1/|\mathscr{X}|$ to each element of $\mathscr{X}$, is also weakly permutation invariant. We shall also see in Examples 5 and 6 of Section 9 that there may be many more coherent lower previsions that share the same weak permutation invariance property. If, however, we strengthen the Symmetry Principle to require weak invariance with respect to *all transformations*, and not just all permutations, then Theorem 4 tells us that the vacuous lower prevision $\underline{P}_{\mathscr{X}}$ is indeed the only coherent lower prevision that is compatible with the following

Strong Symmetry Principle (SSP). *If a subject is completely ignorant about the value of a random variable X in $\mathscr{X}$, then her corresponding belief model should be weakly invariant with respect to the monoid $\mathscr{T}_{\mathscr{X}}$ of all transformations of $\mathscr{X}$.*

Walley [1991, Section 5.5.1 and note 7 on p. 526] has shown that for random variables X taking values in a finite set $\mathscr{X}$, the vacuous lower prevision $\underline{P}_{\mathscr{X}}$ is the only coherent lower prevision that is compatible with SP and the so-called[37]

Embedding Principle (EP). *Consider a random variable X, and consider a set of possible values A for X. Then the (lower) probability assigned to the event A, i.e., the lower probability that $X \in A$, should not depend on the set $\mathscr{X}$ of all possible values for X in which A is embedded.*

So under coherence, SSP is equivalent to SP and EP taken together. Under coherence, it is also equivalent to the following rationality principle, as we shall shortly see.

Revised Principle of Insufficient Reason (RPIR). *If you have two* different *gambles f and g on a random variable X that you are completely ignorant about, then if $f \not\geq g$ you have no reason to prefer f to g.*

Indeed, the only coherent belief model that is compatible with this principle, is the vacuous one. We shall argue in terms of real desirability models[38] $\mathscr{R}$ (see Section 2.1). Say that a subject (really) *prefers* f to g whenever $f \neq g$ and $f - g \in \mathscr{R}$, i.e., she accepts to exchange g for f. Then RPIR implies that for all $f \neq 0$, $f \not\geq 0$ implies that $f \notin \mathscr{R}$, or equivalently, by contraposition, that $f \in \mathscr{R}$ implies $f \geq 0$. Hence $\mathscr{R} = \mathscr{C}_+$ is the vacuous belief model.

In summary, we have the following equivalences, under coherence, and the only belief model that is compatible with these three equivalent rationality requirements, is the vacuous one:

$$\text{SSP} \Leftrightarrow \text{SP+EP} \Leftrightarrow \text{RPIR}.$$

RPIR is a revised version of the Principle of Insufficient Reason (PIR), which states that if you are completely ignorant about the value of a random variable X, then you have no reason to distinguish between the different possible values, and therefore should consider all these values to have equal probability. Indeed, from a historical point of view, the PIR

---

[37]For additional discussion of this principle, see also Walley [1996b], Walley and Bernard [1999].

[38]A similar argument can be given for almost-desirability models $\mathscr{D}$ and lower previsions $\underline{P}$, using for preference Walley's [1991, Sections 3.7.7–3.7.9] corresponding notion of *strict preference*, which corresponds to the present argument by using $\mathscr{D}_{\underline{P}}^+ \cup \{0\}$ as a coherent set of really desirable gambles.

was used extensively by Laplace (see for instance Howie [2002]) to justify using a uniform probability for modelling complete ignorance.

We are of course aware that our reformulation RPIR of Laplace's PIR is quite unusual and has little or no historical grounds, which is why we refer to it as a *revised*, or perhaps better, improved principle. It might have been preferable to call RPIR the 'Principle of Insufficient Reason to Prefer', but we decided against that for aesthetical reasons.

We think that RPIR is reasonable, but that PIR isn't. Indeed, one of the reasons for the critical attitudes of many researchers towards 'Bayesian methods' and inverse probability in the nineteenth and early twentieth century seem to lie in the indiscriminate use by many of Laplace's PIR in order to obtain uniform prior probabilities that can be plugged into Bayes's formula.[39] And by 'indiscriminate use' we mean precisely the confusion that exists between symmetry of evidence and evidence of symmetry: we have argued that it is only evidence of symmetry that justifies using strongly invariant belief models (and in many cases, such as permutation invariance for finite spaces, strong invariance singles out the uniform probability as the only compatible belief model, see also Section 9). If there is only symmetry of evidence, we should use weakly invariant belief models, and in the special case of complete ignorance, vacuous ones. Of course, as we said in the Introduction and proved in the previous section, for precise previsions (Bayesian belief models) there is no difference between weak and strong invariance, so if you insist on using a Bayesian belief model, symmetry of evidence leads you to a (strongly) invariant one! The problem with the PIR, therefore, is that the belief model is only allowed to be precise: there would be fewer or no difficulties if in its formulation we just replaced 'probability' with 'lower and upper probability', for instance.

6. Weakly invariant lower previsions

Let us now turn to a more involved mathematical study of the invariance of coherent lower previsions. So far, we have only looked at coherent lower previsions that were defined on all gambles. But of course, it will usually happen that our subject specifies a supremum acceptable buying price $\underline{P}(f)$ for only a limited number of gambles f, say those in a subset $\mathscr{K}$ of $\mathscr{L}(\mathscr{X})$. And then we can ask ourselves whether such an assessment can be coherently extended to a weakly, or to a strongly, $\mathscr{T}$-invariant lower prevision on all gambles. We shall address these, and related, problems in this and the following section. Let us begin here with weak invariance. The following definition generalises the already established notion of weak invariance to lower previsions defined on any $\mathscr{T}$-invariant domain, that are not necessarily coherent (they may even incur a sure loss).[40]

Definition 4 (Weak invariance). A lower prevision $\underline{P}$ defined on a set of gambles $\mathscr{K} \subseteq \mathscr{L}(\mathscr{X})$ is called weakly $\mathscr{T}$-*invariant* if

(W1) $T^t f \in \mathscr{K}$ for all f in $\mathscr{K}$ and T in $\mathscr{T}$, i.e., $\mathscr{K}$ is $\mathscr{T}$-invariant;
(W2) $\underline{P}(T^t f) \geq \underline{P}(f)$ for all f in $\mathscr{K}$ and T in $\mathscr{T}$, i.e., all $T\underline{P}$ point-wise dominate $\underline{P}$.

As before, if $\mathscr{T}$ is right-cancellable (and in particular if it is a group), the inequality in the invariance definition is actually an equality: consider a gamble f in $\mathscr{K}$, a transformation T in $\mathscr{T}$ and its right-inverse R, we have $\underline{P}(f) = \underline{P}((TR)^t f) = \underline{P}(R^t(T^t f)) \geq \underline{P}(T^t f)$ in addition to $\underline{P}(T^t f) \geq \underline{P}(f)$.

---

[39]An interesting historical discussion of such attitudes can be found in Howie [2002] and Zabell [1989b].

[40]Our notion of weak invariance for a lower prevision is weaker than Walley's [1991, Section 3.5.1] corresponding notion of invariance, which requires equality, and has the drawback that it is not preserved by natural extension.

Next, because taking convex combinations, lower envelopes, limits inferior and superior preserves inequalities, it is easy to see that convex combinations, lower envelopes and point-wise limits of weakly invariant lower previsions are also weakly invariant. Observe by the way that the same operations also preserve coherence.

The following proposition looks at weak invariance for (precise) previsions.

Proposition 7. *Let P be a prevision, i.e., a self-conjugate lower prevision, defined on a negation-invariant domain $\mathscr{K} = -\mathscr{K}$. Assume that $\mathscr{K}$ is also $\mathscr{T}$-invariant. Then P is weakly $\mathscr{T}$-invariant if and only if $P(T^t f) = P(f)$ for all T in $\mathscr{T}$ and all f in $\mathscr{K}$.*

Proof. It is clear that the condition is sufficient. To show that it is also necessary, assume that P is $\mathscr{T}$-invariant, and consider any T in $\mathscr{T}$ and any gamble f in $\mathscr{K}$. Then it follows from the $\mathscr{T}$-invariance of P that on the one hand $P(T^t f) \geq P(f)$, and on the other hand, since $-f \in \mathscr{K}$ and $T^t(-f) = -T^t f \in \mathscr{K}$, that $P(-T^t f) = P(T^t(-f)) \geq P(-f)$, or equivalently, using the self-conjugacy of P, that $P(f) \geq P(T^t f)$. $\qquad\square$

We study next whether a weakly invariant lower prevision $\underline{P}$ with domain $\mathscr{K}$ can be extended to a coherent weakly invariant lower prevision on the set of all gambles, or more generally, whether there is a coherent weakly invariant lower prevision on all gambles that dominates $\underline{P}$. We already know from the material in Section 2.3 that a necessary condition for this is that $\underline{P}$ should avoid sure loss. Indeed, if $\underline{P}$ incurs sure loss then it has no dominating coherent lower prevision, let alone a weakly invariant one. The perhaps surprising result we prove next is that avoiding sure loss is also sufficient, and that all we have to do is consider the natural extension $\underline{E}_P$ of $\underline{P}$, as it preserves weak invariance. This natural extension is automatically guaranteed to be the point-wise smallest weakly $\mathscr{T}$-invariant coherent lower prevision that dominates $\underline{P}$.[41]

Theorem 8 (Natural extension preserves weak invariance). *The natural extension $\underline{E}_P$ of a weakly $\mathscr{T}$-invariant lower prevision $\underline{P}$ on a set of gambles $\mathscr{K}$ that avoids sure loss is still weakly $\mathscr{T}$-invariant, i.e., for all gambles f on $\mathscr{X}$ and all T in $\mathscr{T}$,*

$$T\underline{E}_P(f) = \underline{E}_P(T^t f) \geq \underline{E}_P(f).$$

Consequently, $\underline{E}_P$ is the point-wise smallest weakly $\mathscr{T}$-invariant coherent lower prevision on $\mathscr{L}$ that dominates $\underline{P}$ on its domain $\mathscr{K}$.

Proof. Consider any gamble f on $\mathscr{X}$ and any T in $\mathscr{T}$. From the definition (LPR-NE) of natural extension, and the fact that $T^t \mathscr{K} \subseteq \mathscr{K}$, we get

$$\underline{E}_P(T^t f) = \sup_{\substack{\lambda_k \geq 0, f_k \in \mathscr{K} \\ k=1\ldots,n,n\geq 0}} \left\{ \alpha : T^t f - \alpha \geq \sum_{k=1}^{n} \lambda_k \left[f_k - \underline{P}(f_k) \right] \right\}$$

$$\geq \sup_{\substack{\lambda_k \geq 0, g_k \in \mathscr{K} \\ k=1\ldots,n,n\geq 0,}} \left\{ \alpha : T^t f - \alpha \geq \sum_{k=1}^{n} \lambda_k \left[T^t g_k - \underline{P}(T^t g_k) \right] \right\}. \qquad (7)$$

Now it follows from the T-invariance of $\underline{P}$ that $\underline{P}(T^t g_k) \geq \underline{P}(g_k)$, whence

$$\sum_{k=1}^{n} \lambda_k \left[T^t g_k - \underline{P}(T^t g_k) \right] \leq T^t \sum_{k=1}^{n} \lambda_k \left[g_k - \underline{P}(g_k) \right],$$

---

[41]This result is mentioned, with only a hint at the proof, by Walley [1991, Theorem 3.5.2].

and consequently $f - \alpha \geq \sum_{k=1}^{n} \lambda_k [g_k - \underline{P}(g_k)]$ implies that

$$T^t f - \alpha \geq T^t \sum_{k=1}^{n} \lambda_k [g_k - \underline{P}(g_k)] \geq \sum_{k=1}^{n} \lambda_k \left[T^t g_k - \underline{P}(T^t g_k) \right].$$

So we may infer from the inequality (7) that

$$\underline{E_P}(T^t f) \geq \sup_{\substack{\lambda_k \geq 0, g_k \in \mathscr{K} \\ k=1\ldots,n, n \geq 0,}} \left\{ \alpha : f - \alpha \geq \sum_{k=1}^{n} \lambda_k [g_k - \underline{P}(g_k)] \right\} = \underline{E_P}(f),$$

which completes the proof.   □

Hence, if we start out with a lower prevision $\underline{P}$ on $\mathscr{K}$ that is weakly $\mathscr{T}$-invariant and already coherent, then its natural extension $\underline{E_P}$ is the smallest coherent and weakly $\mathscr{T}$-invariant lower prevision on all gambles that agrees with $\underline{P}$ on $\mathscr{K}$. As we shall show further on, this result does not carry over to strong invariance.

7. STRONGLY INVARIANT LOWER PREVISIONS

We now turn to the study of strong invariance for lower previsions on general domains.

7.1. **Definition and immediate properties.** The following definition generalises the notion of strong invariance introduced in Section 4.2 to lower previsions that needn't be coherent, nor defined on all of $\mathscr{L}(\mathscr{X})$.

Definition 5 (Strong invariance). A lower prevision $\underline{P}$ defined on a set of gambles $\mathscr{K} \subseteq \mathscr{L}(\mathscr{X})$ is called *strongly $\mathscr{T}$-invariant* if

(S1) $T^t f - f \in \mathscr{K}$ and $f - T^t f \in \mathscr{K}$ for all f in $\mathscr{K}$ and all $T \in \mathscr{T}$;

(S2) $\underline{P}(T^t f - f) \geq 0$ and $\underline{P}(f - T^t f) \geq 0$ for all f in $\mathscr{K}$ and all $T \in \mathscr{T}$.

As is the case for weak invariance, it is easy to see that strong $\mathscr{T}$-invariance is preserved under convex combinations, lower envelopes, and point-wise limits, simply because all these operations preserve inequalities.

Proposition 9. *A strongly $\mathscr{T}$-invariant coherent lower prevision on a $\mathscr{T}$-invariant domain is also weakly $\mathscr{T}$-invariant.*

Proof. First of all, the coherence and strong invariance of $\underline{P}$ imply that $0 \leq \underline{P}(T^t f - f) \leq \underline{P}(T^t f) - \underline{P}(f)$, whence $\underline{P}(T^t f) \geq \underline{P}(f)$ and similarly, we derive from $\underline{P}(f - T^t f) \geq 0$ that $\underline{P}(f) \geq \underline{P}(T^t f)$. So we see that $\underline{P}$ is also weakly $\mathscr{T}$-invariant (with equality).   □

To see that a converse result does not generally hold, so weak invariance is actually weaker than strong invariance, consider the vacuous lower prevision $\underline{P}_{\mathscr{X}}$ on $\mathscr{L}(\mathscr{X})$ and the transformation T_{x_0} that maps all elements x of $\mathscr{X}$ to x_0. Then, for any gamble f such that $\inf f < f(x_0)$ we have $\underline{P}_{\mathscr{X}}(f - T_{x_0}^t f) < 0$. Hence, $\underline{P}_{\mathscr{X}}$ is not strongly T_{x_0}-invariant but Theorem 4 implies that it is weakly T_{x_0}-invariant. If we consider a finite space $\mathscr{X}$ and the vacuous lower prevision $\underline{P}_{\mathscr{X}}$ on $\mathscr{L}(\mathscr{X})$ and the class $\mathscr{P}_{\mathscr{X}}$ of all permutations of $\mathscr{X}$, we can see that weak invariance (with equality) does not imply strong invariance.

So weak invariance is indeed a weaker notion than strong invariance. The following theorem expresses the main difference between these two concepts: while the former means that the set of coherent previsions $\mathscr{M}(\underline{P})$ is invariant, the latter means that every element of this set is invariant.

Theorem 10. *Let $\mathscr{K}$ be a negation invariant and $\mathscr{T}$-invariant set of gambles such that $T^t f - f$ is in $\mathscr{K}$ for all f in $\mathscr{K}$ and T in $\mathscr{T}$.*

1. *A coherent prevision P on $\mathscr{K}$ is weakly $\mathscr{T}$-invariant if and only if it is strongly $\mathscr{T}$-invariant. In either case we simply call it $\mathscr{T}$-invariant.*
2. *A coherent lower prevision $\underline{P}$ on $\mathscr{K}$ is strongly $\mathscr{T}$-invariant if and only if all its dominating coherent previsions are (strongly) $\mathscr{T}$-invariant on $\mathscr{K}$.*

Proof. We start with the first statement. We only need to prove the direct implication, so assume that P is weakly $\mathscr{T}$-invariant, and consider any f in $\mathscr{K}$. Then from the assumption and Proposition 7 we get $P(T^t f) = P(f)$, and it follows from the linearity of P that indeed $P(T^t f - f) = P(f - T^t f) = 0$.

We now turn to the second statement. Since any coherent lower prevision is the lower envelope of its dominating coherent previsions, the converse implications follow at once, since taking a lower envelope preserves strong invariance. To prove the direct implication, assume that $\underline{P}$ is strongly $\mathscr{T}$-invariant, and consider any coherent prevision P in $\mathscr{M}(\underline{P})$. For any T in $\mathscr{T}$ and any f in $\mathscr{K}$ we then find that

$$0 \leq \underline{P}(f - T^t f) \leq P(f - T^t f) = -P(T^t f - f) \leq -\underline{P}(T^t f - f) \leq 0,$$

whence indeed $P(f) = P(T^t f)$. $\qquad\qquad\square$

7.2. Strongly invariant natural extension.

We have shown when studying weak invariance that for any weakly $\mathscr{T}$-invariant lower prevision $\underline{P}$ on some domain $\mathscr{K}$ that avoids sure loss, there is a point-wise smallest weakly invariant coherent lower prevision defined on all gambles that dominates it: its natural extension $\underline{E}_P$. Let us now investigate whether something similar can be done for the notion of strong invariance. The question then is: Consider a monoid $\mathscr{T}$ of transformations of $\mathscr{X}$ and a lower prevision $\underline{P}$ on $\mathscr{K}$ that avoids sure loss, are there strongly $\mathscr{T}$-invariant coherent lower previsions on all $\mathscr{L}(\mathscr{X})$ that dominate $\underline{P}$, and if so, what is the point-wise smallest such lower prevision? Let us denote, as before, by

$$\mathscr{D}_{\underline{P}} = \left\{ f \in \mathscr{L}(\mathscr{X}) \colon \underline{E}_P(f) \geq 0 \right\}$$

the set of almost-desirable gambles associated with $\underline{P}$, and by

$$\mathscr{M}(\underline{P}) = \{ P \in \mathbb{P}(\mathscr{X}) \colon (\forall f \in \mathscr{K})(P(f) \geq \underline{P}(f)) \}$$

its set of dominating coherent previsions, then clearly a coherent lower prevision $\underline{Q}$ on $\mathscr{L}(\mathscr{X})$ is strongly $\mathscr{T}$-invariant and dominates $\underline{P}$ if and only if $\mathscr{M}(\underline{Q}) \subseteq \mathscr{M}(\underline{P}) \cap \mathscr{M}(\overline{\mathscr{D}}_{\mathscr{T}})$, or equivalently, $\mathscr{D}_{\underline{P}} \cup \mathscr{D}_{\mathscr{T}} \subseteq \mathscr{D}_{\underline{Q}}$. So there are strongly $\mathscr{T}$-invariant coherent (lower) previsions that dominate $\underline{P}$ if and only if $\mathscr{M}(\underline{P}) \cap \mathscr{M}(\mathscr{D}_{\mathscr{T}}) \neq \emptyset$, or equivalently, if the set of almost-desirable gambles $\mathscr{D}_{\underline{P}} \cup \mathscr{D}_{\mathscr{T}}$ avoids sure loss, and in this case the lower envelope of $\mathscr{M}(\underline{P}) \cap \mathscr{M}(\mathscr{D}_{\mathscr{T}})$, or equivalently, the lower prevision associated with the natural extension of the set of almost-desirable gambles $\mathscr{D}_{\underline{P}} \cup \mathscr{D}_{\mathscr{T}}$, is the smallest such lower prevision. In the language of coherent lower previsions, this leads to the following theorem.[42]

Theorem 11 (Strongly invariant natural extension). *Consider a lower prevision $\underline{P}$ on $\mathscr{K}$ that avoids sure loss, and a monoid $\mathscr{T}$ of transformations of $\mathscr{X}$. Then there are strongly $\mathscr{T}$-invariant coherent (lower) previsions on $\mathscr{L}(\mathscr{X})$ that dominate $\underline{P}$ on $\mathscr{K}$ if and only if*

$$\overline{E}_{\underline{P}} \left(\sum_{k=1}^{n} \left[f_k - T_k^t f_k \right] \right) \geq 0 \quad \text{for all } n \geq 0, f_1, \ldots, f_n \text{ in } \mathscr{L}(\mathscr{X}) \text{ and } T_1, \ldots, T_n \text{ in } \mathscr{T}, \quad (8)$$

---

[42]Walley [1991, Theorems 3.5.2 and 3.5.3] proves similar results involving Eqs. (8) and (10) for what we call weakly $\mathscr{T}$-invariant $\underline{P}$ that avoid sure loss, in a different manner. See also footnotes 33 and 36.

or equivalently, if

$$\overline{E}_{\mathscr{T}}\left(\sum_{k=1}^{n} \lambda_k \left[f_k - \underline{P}(f_k)\right]\right) \geq 0 \quad \text{for all } n \geq 0, \text{ and } f_1, \ldots, f_n \text{ in } \mathscr{K}. \tag{9}$$

In that case the smallest coherent and strongly T-invariant lower prevision on $\mathscr{L}(\mathscr{X})$ that dominates $\underline{P}$ on its domain $\mathscr{K}$ is given by

$$\underline{E}_{\underline{P},\mathscr{T}}(f) = \sup\left\{\underline{E}_{\underline{P}}\left(f - \sum_{k=1}^{n}\left[f_k - T_k^t f_k\right]\right) : n \geq 0, f_k \in \mathscr{L}(\mathscr{X}), T_k \in \mathscr{T}\right\} \tag{10}$$

$$= \sup\left\{\underline{E}_{\mathscr{T}}\left(f - \sum_{k=1}^{n} \lambda_k \left[f_k - \underline{P}(f_k)\right]\right) : n \geq 0, f_k \in \mathscr{K}, \lambda_k \geq 0\right\} \tag{11}$$

for all gambles f on $\mathscr{X}$; and $\mathscr{M}(\underline{E}_{\underline{P},\mathscr{T}})$ is the set of all $\mathscr{T}$-invariant coherent previsions that dominate $\underline{P}$ on $\mathscr{K}$.

Proof. We already know that there is a dominating coherent (lower) prevision if and only if $\mathscr{M}(\underline{P}) \cap \mathscr{M}(\mathscr{D}_{\mathscr{T}})$ is non-empty. Let us show that this is equivalent to the conditions (8) and (9). To see the equivalence between these two conditions, it suffices to notice [use Eq. (LPR-NE) and the fact that $\overline{E}_{\underline{P}}(h) = -\underline{E}_{\underline{P}}(-h)$] that condition (8) is equivalent to

$$\sup\left[\sum_{k=1}^{n}\left[f_k - T_k^t f_k\right] + \sum_{j=1}^{m}\left[g_j - \underline{P}(g_j)\right]\right] \geq 0$$

$$\text{for all } n,m \geq 0, f_k \in \mathscr{L}(\mathscr{X}), T_k \in \mathscr{T}, g_j \in \mathscr{K}, \quad (12)$$

and that this is in turn [use Eq. (6) and the fact that $\overline{E}_{\mathscr{T}}(h) = -\underline{E}_{\mathscr{T}}(-h)$] equivalent to condition (9). But, considering condition (M-ASL), we see that condition (12) holds if and only if the set of almost-desirable gambles $\mathscr{D}_{\underline{P}} \cup \mathscr{D}_{\mathscr{T}}$ avoids sure loss, or equivalently, if the corresponding set of coherent previsions $\mathscr{M}(\underline{P}) \cap \mathscr{M}(\mathscr{D}_{\mathscr{T}})$ is non-empty.

We now prove the validity of the expression (11) for the lower envelope $\underline{E}_{\underline{P},\mathscr{T}}$ of the set of coherent previsions $\mathscr{M}(\underline{P}) \cap \mathscr{M}(\mathscr{D}_{\mathscr{T}})$. The proof for the expression (10) is analogous. We know from the material in Section 2 that this lower envelope is also the coherent lower prevision associated with the natural extension of the set of almost-desirable gambles $\mathscr{D}_{\underline{P}} \cup \mathscr{D}_{\mathscr{T}}$, so we get by applying Eq. (LPR-NE) with $\mathscr{D} = \mathscr{D}_{\underline{P}} \cup \mathscr{D}_{\mathscr{T}}$ that

$$\underline{E}_{\underline{P},\mathscr{T}}(f) = \sup_{\substack{\lambda_k \geq 0, g_k \in \mathscr{D}_{\underline{P}} \\ k=1,\ldots,n,n \geq 0}} \sup_{\substack{\mu_\ell \geq 0, h_\ell \in \mathscr{D}_{\mathscr{T}} \\ \ell=1,\ldots,m,m \geq 0}} \inf\left[f - \sum_{k=1}^{n} \lambda_k g_k - \sum_{\ell=1}^{m} \mu_\ell h_\ell\right]$$

$$= \sup_{\substack{\lambda_k \geq 0, g_k \in \mathscr{D}_{\underline{P}} \\ k=1,\ldots,n,n \geq 0}} \sup_{\substack{\mu_\ell \geq 0, h_\ell \in \mathscr{D}_{\mathscr{T}} \\ \ell=1,\ldots,m,m \geq 0}} \inf\left[\left(f - \sum_{k=1}^{n} \lambda_k g_k\right) - \sum_{\ell=1}^{m} \mu_\ell h_\ell\right]$$

$$= \sup_{\substack{\lambda_k \geq 0, g_k \in \mathscr{D}_{\underline{P}} \\ k=1,\ldots,n,n \geq 0}} \underline{E}_{\mathscr{T}}\left(f - \sum_{k=1}^{n} \lambda_k g_k\right) = \sup_{\substack{\lambda_k \geq 0, f_k \in \mathscr{K} \\ k=1,\ldots,n,n \geq 0}} \underline{E}_{\mathscr{T}}\left(f - \sum_{k=1}^{n} \lambda_k[f_k - \underline{P}(f_k)]\right),$$

for every gamble f on $\mathscr{X}$, also taking into account the definition (6) of $\underline{E}_{\mathscr{T}}$. $\qquad\square$

In conclusion, whenever the equivalent conditions (8) and (9) are satisfied for a lower prevision $\underline{P}$ that avoids sure loss, then (and only then) the functional $\underline{E}_{\underline{P},\mathscr{T}}$, defined by Eqs. (10) and (11), is the point-wise smallest coherent and strongly $\mathscr{T}$-invariant lower

prevision that dominates $\underline{P}$. We shall call $\underline{E}_{P,\mathscr{T}}$ the *strongly $\mathscr{T}$-invariant natural extension* of $\underline{P}$, as it is the belief model that the assessments captured in $\underline{P}$ lead to if in addition a (so-called structural)[43] assessment of symmetry involving the monoid $\mathscr{T}$ is made.

7.3. **The existence of strongly invariant coherent (lower) previsions.** There is a beautiful and surprisingly simple argument to show that for some types of monoids $\mathscr{T}$, there always are strongly $\mathscr{T}$-invariant lower previsions that dominate a given lower prevision that is weakly $\mathscr{T}$-invariant and avoids sure loss. It is based on the combination of a number of ideas in the literature: (i) Agnew and Morse [1938, Section 2] constructed some specific type of Minkowski functional and used this together with a Hahn–Banach extension result to prove the existence of linear functionals that are invariant with respect to certain groups of permutations; (ii) Day [1942, Theorem 3] showed, in a discussion of ergodic theorems, that a similar construction always works for Abelian semigroups of transformations; (iii) with crucially important insight, Walley [1991, Theorems 3.5.2 and 3.5.3] recognised that the Minkowski functional in the existence proofs of Agnew and Morse, and Day, is actually what we have called a strongly invariant lower prevision, and he used the ideas behind this construction to introduce what we shall call *mixture lower previsions* in Section 7.4; (iv) in another seminal discussion of mean ergodic theorems, Alaoglu and Birkhoff [1940] show that (Moore–Smith-like) convergence of convex mixtures of linear transformations is instrumental in characterising ergodicity; and (v) Bhaskara Rao and Bhaskara Rao [1983, Section 2.1.3] use so-called Banach limits to generate shift-invariant probability charges. In this and the next section, we combine and extend these ideas to prove more general existence results for (strongly) invariant coherent (lower) previsions, and to investigate their relation to (generalised) Banach limits (Section 8). As we shall see in Section 7.4, Walley's [1991, Section 3.5] results can then be derived from our more general treatment.

Consider a monoid $\mathscr{T}$ of transformations of $\mathscr{X}$. We can, as before, consider the set of lifted transformations $\mathscr{T}^t$ as a monoid of linear transformations of the linear space $\mathscr{L}(\mathscr{X})$. A *convex combination T^** of elements of $\mathscr{T}^t$ is a linear transformation of $\mathscr{L}(\mathscr{X})$ of the form

$$T^* = \sum_{k=1}^{n} \lambda_k T_k^t,$$

where $n \geq 1$, $\lambda_1, \ldots, \lambda_n$ are non-negative real numbers that sum to one, and of course $T^* f = \sum_{k=1}^{n} \lambda_k T_k^t f$. We denote by $\mathscr{T}^*$ the set of all convex combinations of elements of $\mathscr{T}^t$. We have of course for any two elements $T_1^* = \sum_{k=1}^{m} \lambda_k U_k^t$ and $T_2^* = \sum_{k=1}^{n} \mu_k V_k^t$ of $\mathscr{T}^*$ that their composition

$$T_2^* T_1^* = \sum_{k=1}^{n} \mu_k V_k^t \left(\sum_{\ell=1}^{m} \lambda_\ell U_\ell^t \right) = \sum_{k=1}^{n} \sum_{\ell=1}^{m} \mu_k \lambda_\ell V_k^t U_\ell^t = \sum_{k=1}^{n} \sum_{\ell=1}^{m} \lambda_\ell \mu_k (U_\ell V_k)^t$$

again belongs to $\mathscr{T}^*$. This implies that $\mathscr{T}^*$ is a monoid of linear transformations of $\mathscr{L}(\mathscr{X})$ as well. We can now introduce invariance definitions involving transformations in $\mathscr{T}^*$ in precisely the same way as we defined them for $\mathscr{T}$ (or actually $\mathscr{T}^t$). We can also define, for any real functional Λ and $T^* \in \mathscr{T}^*$, the transformed functional $T^*\Lambda$ as $\Lambda \circ T^*$. We then have the following result.

Proposition 12. *The following statements hold, where f is a gamble on $\mathscr{X}$, $\mathscr{K}$ is a convex set of gambles on $\mathscr{X}$, and $\underline{P}$ is a coherent lower prevision on $\mathscr{K}$:*

1. f is $\mathscr{T}$-invariant if and only if f is $\mathscr{T}^*$-invariant;

---

[43]Structural assessments are discussed in general in Walley [1991, Chapter 9].

2. *$\mathcal{K}$ is $\mathcal{T}$-invariant if and only if $\mathcal{K}$ is $\mathcal{T}^*$-invariant;*
3. *$\underline{P}$ is weakly $\mathcal{T}$-invariant if and only if $\underline{P}$ is weakly $\mathcal{T}^*$-invariant;*
4. *$\underline{P}$ is strongly $\mathcal{T}$-invariant if and only if $\underline{P}$ is strongly $\mathcal{T}^*$-invariant.*

Proof. It suffices of course to prove the direct implications. Consider an arbitrary $T^* = \sum_k \lambda_k T_k \in \mathcal{T}^*$. For the first statement, let f be $\mathcal{T}$-invariant, then $T^* f = \sum_k \lambda_k T_k^t f = \sum_k \lambda_k f = f$, where the second equality follows from the $\mathcal{T}$-invariance of f. So f is $\mathcal{T}^*$-invariant. For the second statement, let $\mathcal{K}$ be $\mathcal{T}$-invariant and let $f \in \mathcal{K}$, then $T^* f = \sum_k \lambda_k T_k^t f \in \mathcal{K}$, because $T_k^t f \in \mathcal{K}$ for all k by the $\mathcal{T}$-invariance of $\mathcal{K}$ and because $\mathcal{K}$ is convex. So $\mathcal{K}$ is $\mathcal{T}^*$-invariant. For the third statement, assume that $\underline{P}$ is weakly $\mathcal{T}$-invariant. For any $f \in \mathcal{K}$,

$$\underline{P}(T^* f) = \underline{P}\left(\sum_k \lambda_k T_k^t f\right) \geq \sum_k \lambda_k \underline{P}(T_k^t f) \geq \sum_k \lambda_k \underline{P}(f) = \underline{P}(f),$$

where the first inequality follows from the coherence of $\underline{P}$, and the second from the weak $\mathcal{T}$-invariance of $\underline{P}$. Hence $\underline{P}$ is weakly $\mathcal{T}^*$-invariant. For the last statement, assume that $\underline{P}$ is strongly $\mathcal{T}$-invariant. For any $f \in \mathcal{K}$,

$$\underline{P}\left(\sum_k \lambda_k T_k^t f - f\right) = \underline{P}\left(\sum_k \lambda_k (T_k^t f - f)\right) \geq \sum_k \lambda_k \underline{P}(T_k^t f - f) \geq 0,$$

where the first inequality follows from the coherence of $\underline{P}$, and the second from the strong $\mathcal{T}$-invariance of $\underline{P}$. Similarly $\underline{P}(f - \sum_k \lambda_k T_k^t f) \geq 0$. Hence $\underline{P}$ is strongly $\mathcal{T}^*$-invariant. $\square$

We now define the following binary relation $\geq$ on $\mathcal{T}^*$: for T_1^* and T_2^* in $\mathcal{T}^*$ we say that T_2^* *is a successor* of T_1^*, and we write $T_2^* \geq T_1^*$, if and only if there is some T^* in $\mathcal{T}^*$ such that $T_2^* = T^* T_1^*$. Clearly $\geq$ is a reflexive and transitive relation, because $\mathcal{T}^*$ is a monoid. We say that $\mathcal{T}^*$ has the *Moore–Smith property*, or is *directed by $\geq$*, if any two elements of $\mathcal{T}^*$ have a common successor, i.e., for any T_1^* and T_2^* in $\mathcal{T}^*$ there is some T^* in $\mathcal{T}^*$ such that $T^* \geq T_1^*$ and $T^* \geq T_2^*$. It is not difficult to see that if $\mathcal{T}$ is Abelian, or a finite group, then $\mathcal{T}^*$ is directed by the successor relation. This need not hold if $\mathcal{T}$ is an infinite group or a finite monoid, however.

Now, given a *net α* on $\mathcal{T}^*$, i.e., a mapping $\alpha \colon \mathcal{T}^* \to \mathbb{R}$, we can take the *Moore–Smith limit* of α with respect to the directed set $(\mathcal{T}^*, \geq)$ [Moore and Smith, 1922, Section I, p. 103], which, if it exists, is uniquely defined as the real number a such that, for every $\varepsilon > 0$, there is a T_ε^* in $\mathcal{T}^*$, such that $|\alpha(T^*) - a| < \varepsilon$ for all $T^* \geq T_\varepsilon^*$. The Moore–Smith limit a of α is denoted by $\lim_{T^* \in \mathcal{T}^*} \alpha(T^*)$. This limit always exists if α is non-decreasing and bounded from above, or if α is non-increasing and bounded from below.

Theorem 13. *Let $\underline{P}$ be a coherent and weakly $\mathcal{T}$-invariant lower prevision on $\mathcal{L}(\mathcal{X})$, and assume that $\mathcal{T}^*$ has the Moore–Smith property. Then for any gamble f on $\mathcal{X}$ the Moore–Smith limit $\lim_{T^* \in \mathcal{T}^*} \underline{P}(T^* f)$ converges to a real number $\underline{Q}_{\underline{P}, \mathcal{T}}(f)$. Moreover, $\underline{Q}_{\underline{P}, \mathcal{T}}$ is the point-wise smallest strongly $\mathcal{T}$-invariant coherent lower prevision on $\mathcal{L}(\mathcal{X})$ that dominates $\underline{P}$ on $\mathcal{L}(\mathcal{X})$, and*

$$\underline{Q}_{\underline{P}, \mathcal{T}}(f) = \sup\{\underline{P}(T^* f) : T^* \in \mathcal{T}^*\} = \sup\left\{\underline{P}\left(\frac{1}{n}\sum_{k=1}^n T_k^t f\right) : n \geq 1, T_1, \ldots, T_n \in \mathcal{T}\right\}.$$

$$(13)$$

Proof. First, fix f in $\mathscr{L}(\mathscr{X})$. Consider T_1^* and T_2^* in $\mathscr{T}^*$, and assume that $T_2^* \geqslant T_1^*$. This means that there is some T^* in $\mathscr{T}^*$ such that $T_2^* = T^* T_1^*$, and consequently we find that

$$\underline{P}(T_2^* f) = \underline{P}(T^*(T_1^* f)) \geq \underline{P}(T_1^* f),$$

where the inequality follows from the fact that $\underline{P}$ is in particular weakly T^*-invariant [observe that $\mathscr{L}(\mathscr{X})$ is convex and that $\underline{P}$ is weakly $\mathscr{T}$-invariant, and apply Proposition 12]. This means that the net $\underline{P}(T^* f)$, $T^* \in \mathscr{T}^*$ is non-decreasing. Since this net is moreover bounded from above [by $\sup f$, since $\underline{P}$ is coherent], it converges to a real number $\underline{Q}_{\underline{P},\mathscr{T}}(f)$, and clearly

$$\underline{Q}_{\underline{P},\mathscr{T}}(f) = \lim_{T^* \in \mathscr{T}^*} \underline{P}(T^* f) = \sup \left\{ \underline{P}(T^* f) \colon T^* \in \mathscr{T}^* \right\}. \tag{14}$$

This tells us that the net of coherent lower previsions $T^* \underline{P}$, $T^* \in \mathscr{T}^*$ converges point-wise to the lower prevision $\underline{Q}_{\underline{P},\mathscr{T}}$, so $\underline{Q}_{\underline{P},\mathscr{T}}$ is a coherent lower prevision as well [taking a point-wise limit preserves coherence]. Since $\mathrm{id}_{\mathscr{X}}^t \in \mathscr{T}^*$, it follows from Eq. (14) that $\underline{Q}_{\underline{P},\mathscr{T}}(f) \geq \underline{P}(\mathrm{id}_{\mathscr{X}}^t f) = \underline{P}(f)$, so $\underline{Q}_{\underline{P},\mathscr{T}}$ dominates $\underline{P}$ on $\mathscr{L}(\mathscr{X})$. We now show that $\underline{Q}_{\underline{P},\mathscr{T}}$ is strongly $\mathscr{T}$-invariant.[44] Consider any f in $\mathscr{L}(\mathscr{X})$ and T in $\mathscr{T}$. Then for any $n \geq 1$, $T_n^* := \frac{1}{n} \sum_{k=1}^n (T^k)^t$ belongs to $\mathscr{T}^*$, and it follows from the coherence of $\underline{P}$ that

$$\underline{P}(T_n^*(f - T^t f)) = \frac{1}{n} \underline{P}(T^t f - (T^{n+1})^t f) \geq \frac{1}{n} \inf \left[T^t f - (T^{n+1})^t f \right]$$
$$= -\frac{1}{n} \sup \left[(T^{n+1})^t f - T^t f \right] \geq -\frac{2}{n} \sup |f|,$$

and consequently

$$\underline{Q}_{\underline{P},\mathscr{T}}(f - T^t f) \geq \sup \left\{ -\frac{2}{n} \sup |f| \colon n \geq 1 \right\} = 0.$$

A similar argument can be given for $\underline{Q}_{\underline{P},\mathscr{T}}(T^t f - f) \geq 0$, so $\underline{Q}_{\underline{P},\mathscr{T}}$ is indeed strongly $\mathscr{T}$-invariant.

Next, consider any strongly $\mathscr{T}$-invariant and coherent lower prevision $\underline{Q}$ on $\mathscr{L}(\mathscr{X})$, and assume that it dominates $\underline{P}$. Then we get for any gamble f on $\mathscr{X}$ and any T^* in $\mathscr{T}^*$:

$$\underline{Q}(f) = \underline{Q}(f - T^* f + T^* f) \geq \underline{Q}(f - T^* f) + \underline{Q}(T^* f) \geq \underline{Q}(T^* f) \geq \underline{P}(T^* f),$$

where the first inequality follows from the coherence of $\underline{Q}$, the second inequality from its strong $\mathscr{T}$-invariance [use Proposition 12], and the last inequality from the fact that $\underline{Q}$ dominates $\underline{P}$. We then deduce from Eq. (14) that $\underline{Q}$ dominates $\underline{Q}_{\underline{P},\mathscr{T}}$. So $\underline{Q}_{\underline{P},\mathscr{T}}$ is indeed the point-wise smallest strongly $\mathscr{T}$-invariant coherent lower prevision on $\mathscr{L}(\mathscr{X})$ that dominates $\underline{P}$ on $\mathscr{L}(\mathscr{X})$.

Finally, let us prove the second equality in Eq. (13). Consider a gamble f and any $\varepsilon > 0$. Then, by Eq. (14), there is some T^* in $\mathscr{T}^*$ such that $\underline{Q}_{\underline{P},\mathscr{T}}(f) \leq \underline{P}(T^* f) + \frac{\varepsilon}{2}$. For this T^*, there are $n \geq 1$, $T_1, \ldots, T_n$ in $\mathscr{T}$ and $\lambda_1, \ldots, \lambda_n \geq 0$ that sum to one, such that $T^* = \sum_{k=1}^n \lambda_k T_k^t$. Let $\rho_1, \ldots, \rho_n$ be non-negative rational numbers satisfying $|\rho_i - \lambda_i| \leq \frac{\varepsilon}{2n \sup |f|}$

---

[44]The idea for this part of the proof is due to Walley [1991, Point (iv) of the proof of Theorem 3.5.3].

such that moreover $\sum_{i=1}^{n} \rho_i = 1$.[45] Now it follows from the coherence of $\underline{P}$ that

$$\underline{P}(T^* f) = \underline{P}\left(\sum_{i=1}^{n} \lambda_i T_i^t f\right) \leq \underline{P}\left(\sum_{i=1}^{n} \rho_i T_i^t f\right) - \underline{P}\left(\sum_{i=1}^{n} (\rho_i - \lambda_i) T_i^t f\right),$$

and also

$$\underline{P}\left(\sum_{i=1}^{n} (\rho_i - \lambda_i) T_i^t f\right) \geq \sum_{i=1}^{n} \underline{P}((\rho_i - \lambda_i) T_i^t f) \geq \sum_{i=1}^{n} \inf(\rho_i - \lambda_i) T_i^t f$$

$$\geq \sum_{i=1}^{n} -\frac{\varepsilon}{2n \sup|f|} \sup|f| = -\frac{\varepsilon}{2},$$

whence

$$\underline{Q}_{\underline{P}, \mathscr{T}}(f) \leq \underline{P}(T^* f) + \frac{\varepsilon}{2} \leq \underline{P}\left(\sum_{i=1}^{n} \rho_i T_i^t f\right) + \varepsilon,$$

and consequently

$$\underline{Q}_{\underline{P}, \mathscr{T}}(f) = \sup\left\{\underline{P}\left(\sum_{i=1}^{n} \rho_i T_i^t f\right) : n \geq 1, T_1, \ldots, T_n \in \mathscr{T}, \rho_1, \ldots, \rho_n \in \mathbb{Q}^+, \sum_{i=1}^{n} \rho_i = 1\right\},$$

where $\mathbb{Q}^+$ denotes the set of non-negative rational numbers. Now, it is easy to see [just consider the least common multiple of the denominators of $\rho_1, \ldots, \rho_n$] that this supremum coincides with the right-hand side of Eq. (13). $\qquad\square$

This result allows us to establish the following corollary. It gives a sufficient condition for the existence of strongly $\mathscr{T}$-invariant lower previsions dominating a given coherent lower prevision $\underline{P}$. The smallest such lower prevision reflects how initial behavioural dispositions, reflected in $\underline{P}$, are modified (strengthened) to $\underline{E}_{\underline{P}, \mathscr{T}}$ when we add the extra assessment of strong invariance with respect to a monoid $\mathscr{T}$ of transformations.

Corollary 14 (Strongly invariant natural extension). *Let $\mathscr{T}$ be a monoid of transformations of $\mathscr{X}$ and let $\underline{P}$ be a weakly $\mathscr{T}$-invariant lower prevision on some set of gambles $\mathscr{K}$, that avoids sure loss. Assume that $\mathscr{T}^*$ has the Moore–Smith property. Then there are strongly $\mathscr{T}$-invariant coherent lower previsions on $\mathscr{L}(\mathscr{X})$ that dominate $\underline{P}$ on $\mathscr{L}(\mathscr{X})$, and the smallest such lower prevision, which is called the* strongly $\mathscr{T}$-invariant natural extension *of $\underline{P}$, is given by $\underline{E}_{\underline{P}, \mathscr{T}} = \underline{Q}_{\underline{E}_{\underline{P}}, \mathscr{T}}$. Moreover, for every $\mathscr{T}$-invariant gamble f we have that $\underline{E}_{\underline{P}, \mathscr{T}}(f) = \underline{E}_{\underline{P}}(f)$.*

Proof. The first part of the proof follows at once from the observation that a coherent lower prevision $\underline{Q}$ on $\mathscr{L}(\mathscr{X})$ dominates $\underline{P}$ on $\mathscr{K}$ if and only if it dominates $\underline{E}_{\underline{P}}$ on all gambles. For the second part of the proof, simply observe that if f is a $\mathscr{T}$-invariant gamble, then $T^* f = f$ and therefore $\underline{E}_{\underline{P}}(T^* f) = \underline{E}_{\underline{P}}(f)$ for all T^* in $\mathscr{T}^*$. $\qquad\square$

Let us show in particular how this result applies when we consider the monoid $\mathscr{T}_T$ generated by a single transformation T:

---

[45]To see that such rational numbers exist, it suffices to consider non-negative rational numbers $\rho_1, \ldots, \rho_{n-1}$ such that $0 \leq \rho_i \leq \lambda_i \leq 1$ and $|\rho_i - \lambda_i| \leq \frac{\varepsilon}{2n^2 \sup|f|}$ for $i = 1, \ldots, n-1$, and to let $\rho_n := 1 - \sum_{i=1}^{n-1} \rho_i \geq 1 - \sum_{i=1}^{n-1} \lambda_i = \lambda_n \geq 0$. Then $\rho_n \in [0,1]$, and for n big enough, and unless we are in the trivial case where $\lambda_i = 1$ for some i, we get $|\rho_n - \lambda_n| \leq \frac{\varepsilon}{2n \sup|f|}$.

Corollary 15. *Let T be a transformation of $\mathscr{X}$ and consider the Abelian monoid $\mathscr{T}_T = \{T^n \colon n \geq 0\}$. Then for any weakly T-invariant lower prevision $\underline{P}$ on some set of gambles $\mathscr{K}$ that avoids sure loss, there are strongly T-invariant coherent (lower) previsions on $\mathscr{L}(\mathscr{X})$ that dominate $\underline{P}$, and the point-wise smallest such lower prevision $\underline{E}_{\underline{P},T}$ is given by*

$$\underline{E}_{\underline{P},T}(f) = \lim_{n \to \infty} \underline{E}_{\underline{P}}\left(\frac{1}{n}\sum_{k=0}^{n-1}(T^k)^t f\right) = \sup_{n \geq 1} \underline{E}_{\underline{P}}\left(\frac{1}{n}\sum_{k=0}^{n-1}(T^k)^t f\right).$$

Proof. The existence of strongly T-invariant coherent (lower) previsions on $\mathscr{L}(\mathscr{X})$ that dominate $\underline{P}$ follows from Corollary 14, and the fact that for any Abelian monoid $\mathscr{T}$, $\mathscr{T}^*$ has the Moore–Smith property. It also follows from this corollary that for any gamble f on $\mathscr{X}$,

$$\underline{E}_{\underline{P},T}(f) = \sup\left\{\underline{E}_{\underline{P}}(T^*f)\colon T^* \in \mathscr{T}_T^*\right\} \geq \sup_{n \geq 1}\underline{E}_{\underline{P}}\left(\frac{1}{n}\sum_{k=0}^{n-1}(T^k)^t f\right).$$

To prove the converse inequality, fix any T^* in $\mathscr{T}_T^*$ and any gamble f on $\mathscr{X}$. Then there is some $N \geq 1$ and non-negative $\lambda_0, \ldots, \lambda_{N-1}$ that sum to one, such that $T^* = \sum_{k=0}^{N-1} \lambda_k (T^k)^t$. Consider the element $S_M^* = \frac{1}{M}\sum_{\ell=0}^{M-1}(T^\ell)^t$ of $\mathscr{T}^*$, where M is any natural number such that $M \geq N$. Observe that

$$S_M^* T^* = \frac{1}{M}\sum_{\ell=0}^{M-1}(T^\ell)^t\left(\sum_{k=0}^{N-1}\lambda_k(T^k)^t\right) = \sum_{\ell=0}^{M-1}\sum_{k=0}^{N-1}\frac{\lambda_k}{M}(T^{k+\ell})^t = \sum_{m=0}^{M+N-2}\mu_m(T^m)^t,$$

where we let, for $0 \leq m \leq M+N-2$,

$$\mu_m := \sum_{k=0}^{N-1}\sum_{\ell=0}^{M-1}\frac{\lambda_k}{M}\delta_{m,k+\ell} = \begin{cases} \sum_{k=0}^{m}\frac{\lambda_k}{M} & \text{if } 0 \leq m \leq N-2 \\ \frac{1}{M} & \text{if } N-1 \leq m \leq M-1 \\ \sum_{k=m-M+1}^{N-1}\frac{\lambda_k}{M} & \text{if } M \leq m \leq M+N-2. \end{cases}$$

This tells us that $\mu_m = \frac{1}{M}$ for $N-1 \leq m \leq M-1$, and $0 \leq \mu_m \leq \frac{1}{M}$ for all other m. If we let $\delta_m := \mu_m - \frac{1}{N+M-1}$, it follows at once that

$$|\delta_m| \leq \begin{cases} \dfrac{N-1}{M(M+N-1)} & \text{if } N-1 \leq m \leq M-1 \\ \dfrac{1}{M+N-1} & \text{if } 0 \leq m \leq N-2 \text{ or } M \leq m \leq M+N-2 \end{cases}$$

Consequently, it follows from the weak $\mathscr{T}$-invariance and the coherence of $\underline{E}_{\underline{P}}$ that

$$\begin{aligned}
\underline{E}_{\underline{P}}(T^*f) & \\
&\leq \underline{E}_{\underline{P}}(S_M^* T^* f) \\
&= \underline{E}_{\underline{P}}\left(S_{M+N-1}^* f + \sum_{m=0}^{M+N-2}\delta_m(T^m)^t f\right) \leq \underline{E}_{\underline{P}}(S_{M+N-1}^* f) + \sum_{m=0}^{M+N-2}|\delta_m|\sup|f| \\
&\leq \underline{E}_{\underline{P}}(S_{M+N-1}^* f) + \sup|f|\left[\frac{N-1}{M(M+N-1)}(M-N+1) + \frac{1}{M+N-1}(2N-2)\right] \\
&= \underline{E}_{\underline{P}}(S_{M+N-1}^* f) + \sup|f|\frac{(N-1)(3M-N+1)}{M(M+N-1)}.
\end{aligned}$$

Recall that f and T^*, and therefore also N are fixed. Consider any $\varepsilon > 0$, then there is some $M_\varepsilon \geq N$ such that $\sup|f|\frac{(N-1)(3M-N+1)}{M(M+N-1)} < \varepsilon$ for all $M \geq M_\varepsilon$, whence

$$\underline{E}_P(T^*f) \leq \underline{E}_P(S^*_{M_\varepsilon+N-1}f) + \varepsilon \leq \sup_{n \geq 1}\underline{E}_P(S^*_n f) + \varepsilon.$$

Since this holds for all $\varepsilon > 0$, we get $\underline{E}_P(T^*f) \leq \sup_{n \geq 1}\underline{E}_P(S^*_n f)$. Taking the supremum over all T^* in $\mathscr{T}^*$ leads to the desired inequality. $\qquad\square$

7.4. **Mixture lower previsions.** The condition established in Theorem 13 is fairly general, and guarantees for instance the existence of $\mathscr{T}$-invariant coherent previsions whenever the monoid $\mathscr{T}$ is Abelian, or a finite group. In case $\mathscr{T}^*$ is not directed, however, as may happen for instance for groups $\mathscr{T}$ that are not finite nor Abelian, there may still be $\mathscr{T}$-invariant coherent previsions, as we shall see in Example 2 below. So we see that the directedness of $\mathscr{T}^*$ is not a necessary condition for the existence of $\mathscr{T}$-invariant coherent previsions.

But consider a weakly $\mathscr{T}$-invariant lower prevision $\underline{P}$ defined on some domain $\mathscr{K}$, that avoids sure loss. Even if $\mathscr{T}^*$ is not directed,[46] we may still associate with $\underline{P}$ a lower prevision $\underline{Q}_{P,\mathscr{T}}$ on $\mathscr{L}(\mathscr{X})$ through Eq. (13):

$$\underline{Q}_{P,\mathscr{T}}(f) = \sup_{T^* \in \mathscr{T}^*} \underline{E}_P(T^*f) = \sup\left\{\underline{E}_P\left(\frac{1}{n}\sum_{k=1}^{n}T_k^t f\right) : n \geq 1, T_1, \ldots, T_n \in \mathscr{T}\right\},$$

where we have replaced the Moore–Smith limit by a supremum (with which it would coincide in case $\mathscr{T}^*$ were directed), and where $\underline{E}_P$ is the natural extension of $\underline{P}$ to all gambles. We shall call this lower prevision the *mixture lower prevision* associated with the weakly invariant $\underline{P}$. The supremum in this expression is finite, since it is dominated by $\sup f$. This mixture lower prevision is not necessarily coherent, but it is still strongly $\mathscr{T}^*$-invariant.[47] Moreover, this mixture lower prevision dominates $\underline{E}_P$, and therefore also $\underline{P}$ [observe that $\underline{E}_P$ is weakly invariant because $\underline{P}$ is]; and if there are $\mathscr{T}$-invariant coherent previsions, it is dominated by the strongly $\mathscr{T}$-invariant natural extension $\underline{E}_{P,\mathscr{T}}$ of $\underline{P}$.[48] This shows that $\mathscr{M}(\underline{Q}_{P,\mathscr{T}}) = \mathscr{M}(\underline{E}_{P,\mathscr{T}})$, since all coherent previsions that dominate the strongly $\mathscr{T}$-invariant $\underline{Q}_{P,\mathscr{T}}$ are necessarily $\mathscr{T}$-invariant. And clearly then, if this mixture lower prevision is coherent, it coincides with the strongly invariant natural extension. So we see that the mixture lower prevision, even if it is not coherent, still allows us to characterise all $\mathscr{T}$-invariant coherent previsions. In particular, there are such invariant coherent previsions if and only if it avoids sure loss.

Example 2 (Directedness is not necessary). Let us consider the space $\mathscr{X}_3 := \{1,2,3\}$, and let T_1 and T_2 be the transformations of $\mathscr{X}$ given by $T_1(1) = 1$, $T_1(2) = 2$, $T_1(3) = 2$ and $T_2(1) = 1$, $T_2(2) = 3$, $T_2(3) = 3$, respectively. Since $T_1T_1 = T_1$, $T_2T_2 = T_2$, $T_2T_1 = T_2$ and $T_1T_2 = T_1$, we deduce that the set of transformations $\mathscr{T} = \{\mathrm{id}_{\mathscr{X}}, T_1, T_2\}$ is a monoid. Let $P_{\{1\}}$ be the coherent prevision on $\mathscr{L}(\mathscr{X})$ given by $P_{\{1\}}(f) = f(1)$ for any gamble f, i.e., all of whose probability mass lies in 1. Then we have $P_{\{1\}}(f) = P_{\{1\}}(T_1^t f) = P_{\{1\}}(T_2^t f)$ for

---

[46]This is the general situation that Walley [1991, Section 3.5] considers, and he doesn't discuss the directedness of $\mathscr{T}^*$. He does consider the special case that $\mathscr{T}$ is Abelian for which he proves that the existence of invariant coherent previsions is guaranteed. The results in this section were first proven by him.

[47]Simply observe that the relevant part (near the end) of the proof of Theorem 13 is not based on the directedness of $\mathscr{T}^*$.

[48]To prove that the mixture lower prevision dominates $\underline{P}$, consider $T^* = \mathrm{id}_{\mathscr{X}}$ in its definition. To prove that it is dominated by the strongly invariant natural extension, take $f_k = f/n$ in the expression (10) for this natural extension.

any gamble f, so $P_{\{1\}}$ is $\mathscr{T}$-invariant. Let us show that $\mathscr{T}^*$ does not have the Moore–Smith property.

Consider T_1^* and T_2^* in $\mathscr{T}^*$ given by $T_1^* = \lambda T_1^t + (1-\lambda)T_2^t$ and $T_2^* = \mu T_1^t + (1-\mu)T_2^t$, with $\lambda \neq \mu$. Let T^* be another element of $\mathscr{T}^*$, so there are non-negative α_1, α_2 and α_3 such that $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and $T^* = \alpha_1 \operatorname{id}_{\mathscr{X}}^t + \alpha_2 T_1^t + \alpha_3 T_2^t$. Now,

$$
\begin{aligned}
T^* T_1^* &= \alpha_1 \lambda \operatorname{id}_{\mathscr{X}}^t T_1^t + \alpha_1 (1-\lambda)\operatorname{id}_{\mathscr{X}}^t T_2^t \\
&\quad + \alpha_2 \lambda T_1^t T_1^t + \alpha_2 (1-\lambda)T_1^t T_2^t + \alpha_3 \lambda T_2^t T_1^t + \alpha_3 (1-\lambda)T_2^t T_2^t \\
&= \alpha_1 \lambda T_1^t + \alpha_1 (1-\lambda)T_2^t + \alpha_2 \lambda T_1^t + \alpha_2 (1-\lambda)T_2^t + \alpha_3 \lambda T_1^t + \alpha_3 (1-\lambda)T_2^t \\
&= \lambda T_1^t + (1-\lambda)T_2^t = T_1^*.
\end{aligned}
$$

Similarly, $T^* T_2^* = T_2^*$ for any $T^* \in \mathscr{T}^*$. This means that T_1^* is the only possible successor of T_1^*, and T_2^* is the only possible successor of T_2^*. Hence, $\mathscr{T}^*$ cannot have the Moore–Smith property. Nevertheless, there is a $\mathscr{T}$-invariant coherent prevision $P_{\{1\}}$.

Let us consider the vacuous, and therefore weakly $\mathscr{T}$-invariant and coherent, lower prevision $\underline{P}_{\mathscr{X}_3}$ on $\mathscr{L}(\mathscr{X}_3)$, and the mixture lower prevision $\underline{Q}_{\underline{P}_{\mathscr{X}_3},\mathscr{T}}$ that corresponds with it. It is easy to show that for any gamble f, $\underline{Q}_{\underline{P}_{\mathscr{X}_3},\mathscr{T}}(f) = \min\{f(1), \max\{f(2), f(3)\}\}$ and this lower prevision avoids sure loss, and is therefore strongly $\mathscr{T}$-invariant, but it is not coherent [it is not super-additive]. It is easy to see that $P_{\{1\}}$ is the only coherent prevision that dominates $\underline{Q}_{\underline{P}_{\mathscr{X}_3},\mathscr{T}}$, and is therefore the only $\mathscr{T}$-invariant coherent prevision. ♦

7.5. **Invariance and Choquet integration.** Until now, we have explored the relation between coherence and (weak or strong) invariance. To complete this section, we intend to explore this relation for the particular case of the *n*-monotone lower previsions and probabilities introduced near the end of Section 2.4.

Consider an *n*-monotone lower probability $\underline{P}$ defined on a lattice of events $\mathscr{K}$ containing $\emptyset$ and $\mathscr{X}$. Then its natural extension to all events coincides with its inner set function $\underline{P}_*$, which is given by $\underline{P}_*(A) = \sup\{\underline{P}(B) \colon B \in \mathscr{K}, B \subseteq A\}$. Furthermore, the natural extension to all gambles is given by the Choquet integral with respect to $\underline{P}_*$:

$$
\underline{E}_{\underline{P}}(f) = (C)\int_{\mathscr{X}} f \, \mathrm{d}\underline{P}_* := \inf f + (R)\int_{\inf f}^{\sup f} \underline{P}_*(\{x \in \mathscr{X} \colon f(x) \geq \alpha\}) \, \mathrm{d}\alpha
$$

for all gambles f on $\mathscr{X}$, where the integral on the right-hand side is a Riemann integral. This natural extension (and therefore also the inner set function) is still *n*-monotone [De Cooman et al., 2005b,a]. Since we have proven in Theorem 8 that natural extension preserves weak invariance, we can deduce that the inner set function of a *n*-monotone weakly invariant coherent lower probability, and the associated Choquet functional, are still weakly invariant, *n*-monotone and coherent. We now show that weak invariance of the inner set function and the associated Choquet integral is still guaranteed if the lower probability $\underline{P}$ is not coherent or 2-monotone, but only monotone. In what follows, it is important to remember that for a transformation T of $\mathscr{X}$ and a subset A of $\mathscr{X}$, $T^t I_A = I_{T^{-1}(A)}$.

Proposition 16. *Let $\underline{P}$ be a weakly $\mathscr{T}$-invariant monotone lower probability, defined on a $\mathscr{T}$-invariant lattice of events $\mathscr{K}$ that contains $\emptyset$ and $\mathscr{X}$, and such that $\underline{P}(\emptyset) = 0$ and $\underline{P}(\mathscr{X}) = 1$. Then*

1. the inner set function $\underline{P}_*$ of $\underline{P}$ is weakly $\mathscr{T}$-invariant; and

2. the Choquet integral with respect to $\underline{P}_*$ is weakly $\mathscr{T}$-invariant.

Proof. To prove the first statement, consider any $A \subseteq \mathscr{X}$, and let $B \in \mathscr{K}$ be a any subset of A. Then for any T in $\mathscr{T}$, $T^{-1}(B) \in \mathscr{K}$ and $T^{-1}(B) = \{x \colon Tx \in B\} \subseteq \{x \colon Tx \in A\} =$

$T^{-1}(A)$, whence $\underline{P}(B) \leq \underline{P}(T^{-1}(B)) \leq \underline{P}_*(T^{-1}(A))$, where the first inequality follows from the weak invariance of $\underline{P}$, and the second from the fact that $\underline{P}_*$ is monotone and coincides with $\underline{P}$ on its domain, because $\underline{P}$ is assumed to be monotone. Consequently $\underline{P}_*(A) = \sup_{B \in \mathscr{K}, B \subseteq A} \underline{P}(B) \leq \underline{P}_*(T^{-1}(A))$. Hence, $\underline{P}_*$ is also weakly $\mathscr{T}$-invariant.

To prove the second statement, let f be any gamble on $\mathscr{X}$. Define, for any α in $\mathbb{R}$, the level set $f_\alpha := \{x: f(x) \geq \alpha\}$. Then by the first statement,

$$\underline{P}_*(f_\alpha) \leq \underline{P}_*(T^{-1}(f_\alpha)) = \underline{P}_*(\{x: Tx \in f_\alpha\}) = \underline{P}_*(\{x: f(Tx) \geq \alpha\}) = \underline{P}_*((T^t f)_\alpha).$$

Hence,

$$(C) \int f \, d\underline{P}_* = \inf f + (R) \int_{\inf f}^{\sup f} \underline{P}_*(f_\alpha) \, d\alpha$$

$$\leq \inf f + (R) \int_{\inf f}^{\sup f} \underline{P}_*((T^t f)_\alpha) \, d\alpha = (C) \int T^t f \, d\underline{P}_*,$$

also taking into account for the last equality that $\underline{P}_*((T^t f)_\alpha) = 1$ for all α in $[\inf f, \inf T^t f)$, and that $\underline{P}_*(f_\alpha) = 0$ for all α in $(\sup T^t f, \sup f]$. $\qquad\square$

As we said before, natural extension does not preserve strong invariance in general, and a simple example shows that this continues to hold in particular for n-monotone lower previsions: the unique coherent lower prevision defined on $\{\emptyset, \mathscr{X}\}$ is trivially completely monotone and strongly invariant with respect to any monoid of transformations $\mathscr{T}$, but its natural extension, the vacuous lower prevision $\underline{P}_\mathscr{X}$ (which is completely monotone), is not strongly $\mathscr{T}$-invariant unless in the trivial case that $\mathscr{T} = \{\mathrm{id}_\mathscr{X}\}$.

It is nonetheless interesting that if we restrict ourselves to coherent previsions (which constitute a particular instance of completely monotone lower previsions), natural extension from events to gambles does preserve strong invariance. This is a consequence of the following theorem.

Theorem 17. *Let $\underline{P}$ be a coherent lower prevision on $\mathscr{L}(\mathscr{X})$ and let $\mathscr{T}$ be a monoid of transformations on $\mathscr{X}$. Then $\underline{P}$ is strongly $\mathscr{T}$-invariant if and only if any P in $\mathscr{M}(\underline{P})$, its restriction to events is (weakly) $\mathscr{T}$-invariant, in the sense that $P(T^{-1}(A)) = P(A)$ for all $A \subseteq \mathscr{X}$ and all $T \in \mathscr{T}$.*

Proof. We start with the direct implication. If $\underline{P}$ is strongly $\mathscr{T}$-invariant, then any P in $\mathscr{M}(\underline{P})$ is $\mathscr{T}$-invariant by Theorem 10. Hence, given $A \subseteq \mathscr{X}$ and $T \in \mathscr{T}$, we get $P(A) = P(T^{-1}(A))$.

Conversely, consider P in $\mathscr{M}(\underline{P})$. Recall that a coherent prevision on all events has only one coherent extension from all events to all gambles, namely its natural extension, or Choquet functional; see [De Cooman et al., 2005a]. So for any gamble f on $\mathscr{X}$ and any T in $\mathscr{T}$, taking into account that P is assumed to be invariant on events, and that $T^{-1}(f_\alpha) = (T^t f)_\alpha$ [see the proof of Proposition 16], we get

$$P(f) = (C) \int_\mathscr{X} f \, dP = \inf f + (R) \int_{\inf f}^{\sup f} P(f_\alpha) \, d\alpha$$

$$= \inf f + (R) \int_{\inf f}^{\sup f} P(T^{-1}(f_\alpha)) \, d\alpha = \inf f + (R) \int_{\inf f}^{\sup f} P((T^t f)_\alpha) \, d\alpha$$

$$= (C) \int_\mathscr{X} T^t f \, dP = P(T^t f).$$

Hence, P is strongly $\mathscr{T}$-invariant and, applying Theorem 10, so is the lower envelope $\underline{P}$ of $\mathscr{M}(\underline{P})$. $\qquad\square$

We see that, although the condition of strong invariance cannot be considered for lower probabilities, in the sense that $I_A - T^t I_A$ will not be in general the indicator of an event, it is still to some extent characterised by behaviour on events. Moreover, we may deduce the following result.

Corollary 18. *Let $\underline{P}$ be a strongly $\mathscr{T}$-invariant lower prevision on a $\mathscr{T}$-invariant set of gambles $\mathscr{K}$ that includes all indicators of events. Assume that $\underline{P}$ avoids sure loss. Then its natural extension to all gambles is strongly $\mathscr{T}$-invariant, and coincides therefore with the strongly invariant natural extension of $\underline{P}$.*

Proof. Since $\underline{P}$ avoids sure loss, $\mathscr{M}(\underline{P})$ is non-empty. Since $\underline{P}$ is strongly invariant on a domain that includes all events, any element P of $\mathscr{M}(\underline{P})$ is (strongly) invariant on all events. Hence, by the previous theorem, P is also (strongly) invariant on all gambles, since a coherent prevision on all events has only one coherent extension from all events to all gambles (namely its natural extension, or Choquet functional). Therefore, the natural extension of $\underline{P}$ is a lower envelope of invariant coherent previsions, and is therefore strongly invariant. $\qquad\square$

This result provides further insight into the existence problem for strongly invariant coherent lower previsions. The existence of strongly invariant coherent lower previsions on all gambles is equivalent to the existence of invariant coherent previsions on all gambles, which in turn is equivalent to the existence of invariant coherent previsions on all events (or in other words, invariant finitely additive probabilities). And it is the impossibility of satisfying invariance with finitely additive probabilities in some cases (for instance for the class $\mathscr{T}_\mathscr{X}$ of all transformations) that prevents the existence of coherent strongly invariant belief models.

We also infer that if the restriction $\underline{Q}$ of a coherent lower prevision $\underline{P}$ on $\mathscr{L}(\mathscr{X})$ to gambles of the type $I_A - T^t I_A$ and $T^t I_A - I_A$, involving only indicators of events, is strongly invariant, then $\underline{P}$ is strongly invariant on all of $\mathscr{L}(\mathscr{X})$: it will dominate the natural extension $\underline{E}_Q$ of $\underline{Q}$, which is strongly invariant by Corollary 18, and consequently it will also be strongly invariant.

We can also deduce the following result. Recall that a linear lattice of gambles $\mathscr{K}$ is a set of gambles that is at once a lattice of gambles and a linear subspace of $\mathscr{L}(\mathscr{X})$. If in addition $\mathscr{K}$ contains all constant gambles, then for any coherent prevision P defined on $\mathscr{K}$, its natural extension to all gambles [Walley, 1991, Theorem 3.1.4] is given by the *inner extension* $P_*(f) := \sup\{P(g)\colon g \in \mathscr{K}, g \le f\}$. Let us denote by P^* the conjugate upper prevision of P_*.

Corollary 19. *Let $\mathscr{T}$ be a monoid of transformations of $\mathscr{X}$, and let $\underline{P}$ be a strongly $\mathscr{T}$-invariant lower prevision on a linear lattice of gambles $\mathscr{K}$ that contains all constant gambles. The natural extension $\underline{E}_P$ of $\underline{P}$ to all gambles is strongly $\mathscr{T}$-invariant if and only if for any coherent prevision P on $\mathscr{K}$ that dominates $\underline{P}$, we have $P_*(A \setminus T^{-1}(A)) = P^*(A \setminus T^{-1}(A)) = P_*(T^{-1}(A) \setminus A) = P^*(T^{-1}(A) \setminus A)$ for all $A \subseteq \mathscr{X}$ and all $T \in \mathscr{T}$.*

Proof. It follows from Walley [1991, Theorem 3.4.2] that $\underline{E}_P$ is the lower envelope of the coherent lower previsions P_*, where P is any coherent prevision on $\mathscr{K}$ that dominates $\underline{P}$ on $\mathscr{K}$. But then, clearly, $\underline{E}_P$ will be strongly $\mathscr{T}$-invariant if and only if all the P_* are. Consider any such P_*. By Theorem 17, P_* is strongly invariant if and only if for all $A \subseteq \mathscr{X}$ and $T \in \mathscr{T}$:

$$Q(A) = Q(T^{-1}(A)) \quad \text{for all } Q \text{ in } \mathscr{M}(P_*)$$

which is obviously equivalent to $P_*(I_A - T^t I_A) = P_*(T^t I_A - I_A) = 0$. Now observe that $I_A - T^t I_A = I_A - I_{T^{-1}(A)} = I_{A \setminus T^{-1}(A)} - I_{T^{-1}(A) \setminus A}$, and that the functions $I_{A \setminus T^{-1}(A)}$ and $-I_{T^{-1}(A) \setminus A}$ are comonotone. Since P is a coherent prevision on $\mathscr{K}$, it is completely monotone. Hence, its inner extension P_* is coherent and completely monotone on all gambles, and therefore comonotone additive [De Cooman et al., 2005a]. This means that

$$P_*(I_A - T^t I_A) = P_*(I_{A \setminus T^{-1}(A)} - I_{T^{-1}(A) \setminus A}) = P_*(I_{A \setminus T^{-1}(A)}) + P_*(-I_{T^{-1}(A) \setminus A})$$
$$= P_*(I_{A \setminus T^{-1}(A)}) - P^*(I_{T^{-1}(A) \setminus A}) = P_*(A \setminus T^{-1}(A)) - P^*(T^{-1}(A) \setminus A)$$

and similarly $P_*(T^t I_A - I_A) = P_*(T^{-1}(A) \setminus A) - P^*(A \setminus T^{-1}(A))$. The rest of the proof is now immediate. $\qquad\square$

8. SHIFT-INVARIANCE AND ITS GENERALISATIONS

8.1. **Strongly shift-invariant coherent lower previsions on** $\mathscr{L}(\mathbb{N})$. Let us consider, as an example, the case of the shift-invariant, i.e., $\mathscr{T}_\theta$-invariant, coherent previsions on $\mathscr{L}(\mathbb{N})$. These are usually called *Banach limits* in the literature, see for instance, Bhaskara Rao and Bhaskara Rao [1983, Section 2.1.3] or Walley [1991, Sections 2.9.5 and 3.5.7]. We know from Corollary 14 that there are always Banach limits that dominate a given weakly shift-invariant lower prevision—so we know that there actually are Banach limits. Let us denote by $\mathbb{P}_\theta(\mathbb{N})$ the set of all Banach limits. We also know that a coherent lower prevision on $\mathscr{L}(\mathbb{N})$ is strongly shift-invariant if and only if it is a lower envelope of such Banach limits. The smallest strongly shift-invariant coherent lower prevision $\underline{E}_\theta$ on $\mathscr{L}(\mathbb{N})$ is the lower envelope of all Banach limits, and it is given by:[49]

$$\underline{E}_\theta(f) = \sup_{\substack{m_1,\dots m_n \geq 0 \\ n \geq 0}} \inf_{k \geq 0} \frac{1}{n} \sum_{\ell=1}^{n} f(k + m_\ell) = \lim_{n \to \infty} \inf_{k \geq 0} \frac{1}{n} \sum_{\ell=k}^{k+n-1} f(\ell), \tag{15}$$

for any gamble f on $\mathbb{N}$ (or in other words, for any bounded sequence $f(n)_{n \in \mathbb{N}}$ of real numbers). The first equality follows from Corollary 14, and the second from Corollary 15. $\underline{E}_\theta(f)$ is obtained by taking the infimum sample mean of f over 'moving windows' of length n, and then letting the window length n go to infinity. Since this is the lower prevision on $\mathscr{L}(\mathbb{N})$ that can be derived *solely* using considerations of coherence and the evidence of shift-invariance, we believe that this $\underline{E}_\theta$ is a natural candidate for a '*uniform distribution*' on $\mathbb{N}$. It is the belief model to use if we only have evidence of shift-invariance, as all other strongly shift-invariant coherent lower previsions will point-wise dominate $\underline{E}_\theta$, and will therefore represent stronger behavioural dispositions than warranted by the mere evidence of shift-invariance.[50]

We could also sample f over the set $\{1, \dots, n\}$ leading to a coherent 'sampling' prevision

$$S_n(f) = \frac{1}{n} \sum_{\ell=0}^{n-1} f(\ell),$$

---

[49]See also Walley [1991, Section 3.5.7]. The expression on the right hand side is not a limit inferior!

[50]But this belief model has the important defect that, like the lower prevision $\underline{S}_\theta$ defined further on, it is not fully conglomerable; see Walley [1991, Section 6.6.7] and observe that the counterexample that Walley gives for $\underline{S}_\theta$, also applies to $\underline{E}_\theta$. Walley's remark there that his example shows that there are no (what we call) fully conglomerable (strongly) shift-invariant (lower) previsions that dominate $\underline{S}_\theta$, can be extended in a straightforward manner to $\underline{E}_\theta$ to show that *there are no fully conglomerable (strongly) shift-invariant (lower) previsions*.

but the problem here is that for any given f the sequence of sampling averages $S_n(f)$ is not guaranteed to converge. Taking the limits inferior of such sequences (one for each gamble f), however, yields a coherent lower prevision[51] $\underline{S}_\theta$ given by

$$\underline{S}_\theta(f) = \liminf_{n\to\infty} S_n(f) = \liminf_{n\to\infty} \frac{1}{n} \sum_{\ell=0}^{n-1} f(\ell)$$

for any gamble f on $\mathbb{N}$. For any event $A \subseteq N$, or equivalently, any zero-one-valued sequence, we have that $S_n(A) = \frac{1}{n}|A \cap \{0,\ldots,n-1\}|$ is the 'relative frequency' of ones in the sequence $I_A(n)$ and

$$\underline{S}_\theta(A) = \liminf_{n\to\infty} S_n(A) = \liminf_{n\to\infty} \frac{1}{n}|A \cap \{0,\ldots,n-1\}|.$$

Let $\overline{S}_\theta$ denote the conjugate of $\underline{S}_\theta$, given by $\overline{S}_\theta(f) = \limsup_n S_n(f)$. Those events A for which $\underline{S}_\theta(A) = \overline{S}_\theta(A)$ have a 'limiting relative frequency' equal to this common value. It is not difficult to show that the coherent 'limiting relative frequency' lower prevision $\underline{S}_\theta$ is actually also strongly shift-invariant.[52] This implies that all the coherent previsions that dominate $\underline{S}_\theta$ are strongly shift-invariant. But it is easy to see (see Example 3 below) that $\underline{E}_\theta$ is strictly dominated by $\underline{S}_\theta$, so there are Banach limits that do not dominate $\underline{S}_\theta$.

Proposition 20. *Let L be any Banach limit on $\mathscr{L}(\mathbb{N})$, let f be any gamble on $\mathbb{N}$. Then the following statements hold.*

1. $\liminf_{n\to\infty} f(n) \leq \underline{E}_\theta(f) \leq \underline{S}_\theta(f) \leq \overline{S}_\theta(f) \leq \overline{E}_\theta(f) \leq \limsup_{n\to\infty} f(n)$.
2. *If $\lim_{n\to\infty} f(n)$ exists, then*

$$\underline{E}_\theta(f) = \underline{S}_\theta(f) = \overline{E}_\theta(f) = \overline{S}_\theta(f) = L(f) = \lim_{n\to\infty} f(n).$$

3. *If f is θ^m-invariant (has period $m \geq 1$), then*

$$\underline{E}_\theta(f) = \underline{S}_\theta(f) = \overline{E}_\theta(f) = \overline{S}_\theta(f) = L(f) = \frac{1}{m} \sum_{r=1}^{m-1} f(r).$$

4. *If f is zero except in a finite number of elements of $\mathbb{N}$, then $\underline{E}_\theta(f) = \underline{S}_\theta(f) = \overline{E}_\theta(f) = \overline{S}_\theta(f) = L(f) = 0$. In particular, this holds for the indicator of any finite subset A of $\mathbb{N}$.*

Proof. We begin with the first statement. By conjugacy, we can concentrate on the lower previsions. We have already argued that $\underline{S}_\theta$ is a strongly shift-invariant coherent lower prevision, so $\underline{S}_\theta$ will dominate the smallest strongly shift-invariant coherent lower prevision $\underline{E}_\theta$. So it remains to prove that $\underline{E}_\theta$ dominates the limit inferior. Consider the first equality in Eq. (15). Fix the natural numbers $n \geq 1$, $m_1, \ldots m_n$. We can assume without loss of generality that the m_1 is the smallest of all the m_ℓ. Observe that

$$\inf_{k\geq 0} \frac{1}{n} \sum_{\ell=1}^{n} f(k+m_\ell) \geq \inf_{k\geq 0} \min_{\ell=1}^{n} f(k+m_\ell) = \min_{\ell=1}^{n} \inf_{k\geq m_\ell} f(k) = \inf_{k\geq m_1} f(k),$$

and therefore

$$\underline{E}_\theta(f) \geq \sup_{m_1 \geq 0} \inf_{k\geq m_1} f(k) = \liminf_{n\to\infty} f(n).$$

---

[51] A limit inferior of a sequence of coherent lower previsions is always coherent, see Walley [1991, Corollary 2.6.7].

[52] The following simple proof is due to Walley [1991, Section 3.5.7]. Observe that $S_n(\theta^t f - f) = [f(n) - f(0)]/n \to 0$ as $n \to \infty$, so $\underline{S}_\theta(\theta^t f - f) = \overline{S}_\theta(\theta^t f - f) = 0$.

The second statement is an immediate consequence of the first, and the third follows easily from the definition of $\underline{E}_\theta$ and $\overline{E}_\theta$. Finally, the fourth statement follows at once from the second. $\qquad\square$

Example 3 (Not all Banach limits dominate $\underline{S}_\theta$). Consider the event

$$A = \left\{n^2 + k \colon n \geq 1, k = 0, \ldots, n-1\right\}.$$

Then A has 'limiting relative frequency' $\underline{S}_\theta(A) = \overline{S}_\theta(A) = 1/2$, whereas $\underline{E}_\theta(A) = 0$ and $\overline{E}_\theta(A) = 1$. This shows that $\underline{S}_\theta$ strictly dominates $\underline{E}_\theta$, so not all Banach limits dominate $\underline{S}_\theta$.

Indeed, for the limiting relative frequency, consider the subsequence $S_{m^2-1}(A)$, $m \geq 2$ of $S_n(A)$. Then

$$S_{m^2-1}(A) = \frac{1}{m^2-1}|A \cap \{0, \ldots, m^2-2\}| = \frac{1+2+\cdots+m-1}{m^2-1} = \frac{\frac{1}{2}m(m-1)}{m^2-1} = \frac{1}{2}\frac{m}{m+1},$$

so this subsequence converges to $\frac{1}{2}$. Now the 'integer intervals' $[m^2-1, (m+1)^2-1]$, $m \geq 1$ cover the set of all natural numbers, and as n varies over such an interval, $S_n(A)$ starts at $S_{m^2-1}(A) = \frac{1}{2}\frac{m}{m+1} < \frac{1}{2}$, increases to $S_{m^2+m}(A) = \frac{1}{2}\frac{m^2+m}{m^2+m} = \frac{1}{2}$, and then again decreases to $S_{(m+1)^2-1}(A) = \frac{1}{2}\frac{m+1}{m+2} < \frac{1}{2}$. Both the lower and upper bounds converge to $\frac{1}{2}$ as $m \to \infty$, and therefore the sequence $S_n(A)$ converges to $\frac{1}{2}$ as well.

To calculate $\underline{E}_\theta(A)$, we consider the second equality in Eq. (15). Fix $n \geq 1$ and let $k = n^2 + n$, then $k + n - 1 = (n+1)^2 - 2$, so

$$\frac{1}{n}\sum_{\ell=k}^{k+n-1} I_A(\ell) = \frac{1}{n}\sum_{\ell=n^2+n}^{(n+1)^2-2} I_A(\ell) = 0,$$

whence $\inf_{k \geq 0} \frac{1}{n}\sum_{\ell=k}^{k+n-1} I_A(\ell) = 0$ for all $n \geq 1$, and therefore $\underline{E}_\theta(A) = 0$. To calculate $\overline{E}_\theta(A)$, fix $n \geq 1$ and let $k = n^2$ then

$$\frac{1}{n}\sum_{\ell=k}^{k+n-1} I_A(\ell) = \frac{1}{n}\sum_{\ell=n^2}^{n^2+n-1} I_A(\ell) = 1,$$

whence $\sup_{k \geq 0} \frac{1}{n}\sum_{\ell=k}^{k+n-1} I_A(\ell) = 1$ for all $n \geq 1$, and therefore $\overline{E}_\theta(A) = 1$. $\blacklozenge$

In an interesting paper, Kadane and O'Hagan [1995] study candidates for the 'uniform distribution' on $\mathbb{N}$. They consider, among others, all the finitely additive probabilities (or equivalently, all coherent previsions) that coincide with the limiting relative frequency on all events for which this limit exists. One could also consider as such candidates the coherent previsions that dominate the sampling lower prevision $\underline{S}_\theta$, which have the benefit of being strongly shift-invariant. But, we actually believe that *all* Banach limits (or actually, their lower envelope) are good candidates for being called 'uniform distributions on $\mathbb{N}$' and not just the ones that dominate $\underline{S}_\theta$. Kadane and O'Hagan also propose to consider other coherent previsions, and their idea is to consider the 'residue sets', which are the subsets

$$R_m^r = \{km + r \colon k \geq 0\} = \{\ell \in \mathbb{N} \colon \ell = r \mod m\}$$

of $\mathbb{N}$, where $m \geq 1$ and $r = 1, \ldots, m-1$. These sets are θ^m-invariant, so we already know from Proposition 20 that $\underline{E}_\theta(R_m^r) = \underline{S}_\theta(R_m^r) = \overline{S}_\theta(R_m^r) = \overline{E}_\theta(R_m^r) = \frac{1}{m}$ for all $m \geq 1$ and $r = 1, \ldots, m-1$. Now what Kadane and O'Hagan do, is consider the set of all coherent previsions (finitely additive probabilities in their paper, but that is equivalent) that extend the probability assessments $P(R_m^r) = 1/m$ for all events R_m^r. In other words, they consider

the natural extension $\underline{E}_{\text{res}}$ of all such assessments, i.e., the lower envelope of all such coherent previsions. It is not difficult to prove that this natural extension is given by[53]

$$\underline{E}_{\text{res}}(f) = \lim_{m \to \infty} \frac{1}{m} \sum_{r=0}^{m-1} \inf_{k \in \mathbb{N}} f(km+r).$$

This coherent lower prevision is completely monotone [as a point-wise limit of completely monotone lower previsions, even (natural extensions to gambles of so-called) belief functions [Shafer, 1976]], and weakly shift-invariant [since the natural extension of any weakly shift-invariant lower prevision is]. Since the assessments $P(R_m^r) = \frac{1}{m}$ coincide with the values given by $\underline{E}_\theta$, we see that $\underline{E}_\theta$ will point-wise dominate the natural extension $\underline{E}_{\text{res}}$ of these assessments to all gambles. But as we shall shortly prove in Example 4, $\underline{E}_{\text{res}}$ is not strongly shift-invariant, meaning that among the coherent previsions that extend these assessments, there also are coherent previsions that are not Banach limits (not shift-invariant).

Example 4. Here we show by means of a counterexample that $\underline{E}_{\text{res}}$ is not strongly shift-invariant. Let $B_m := \{0, \ldots, m-1\}$ and $A := \bigcup_{m \geq 1} \{m\} \times B_m$, and consider the map

$$\phi : A \to \mathbb{N} : (m,r) \mapsto \phi(m,r) := \frac{m(m-1)}{2} + r + 1.$$

It is easy to see that ϕ is a bijection (one-to-one and onto). Also define the map

$$\kappa : A \to \mathbb{N} : (m,r) \mapsto \kappa(m,r) := Nm\phi(m,r) + r.$$

for some fixed $N \geq 2$. We consider the strict order $<$ on A induced by the bijection ϕ, i.e., $(m,r) < (m',r')$ if and only if $\phi(m,r) < \phi(m',r')$ [if and only if $m < m$, or $m = m'$ and $r < r'$, so $<$ is the lexicographic order]. Then κ is an increasing map with respect to this order. To see this, assume that $(m,r) < (m',r')$. If $m < m'$, then

$$\kappa(m,r) = Nm\phi(m,r) + r < Nm\phi(m',0) + r$$
$$< Nm'\phi(m',0) + 0 \leq Nm'\phi(m',r') + r' = \kappa(m,r').$$

If on the other hand $m = m'$ and $r < r'$, then $\kappa(m,r) = Nm\phi(m,r) + r < Nm\phi(m,r') + r' = \kappa(m,r')$.

Moreover, given $(m,r) < (m',r')$, we see that $\kappa(m',r') - \kappa(m,r) \geq N$. Indeed, since κ is increasing, it suffices to prove this for consecutive pairs in the order $<$ we have defined on A. There are only two possible expressions of consecutive pairs (m,r) and (m',r'): either we have $(m',r') = (m,r+1)$, and then we get

$$\kappa(m,r+1) - \kappa(m,r) = Nm[\phi(m,r+1) - \phi(m,r)] + 1 = Nm + 1 \geq N;$$

or we have $r = m-1, (m',r') = (m+1,0)$, and then we get

$$\kappa(m+1,0) - \kappa(m,m-1) = Nm[\phi(m+1,0) - \phi(m,m-1)] + N\phi(m+1,0) - (m-1)$$
$$= Nm + N\phi(m+1,0) - (m-1) \geq Nm \geq N,$$

taking into account that $\phi(m+1,0) \geq m-1$ by definition of ϕ.

Consider the set $C = \kappa(A)^c$. Then $\underline{E}_{\text{res}}(C) = \lim_{m \to \infty} \frac{1}{m} \sum_{r=0}^{m-1} \inf_{k \in \mathbb{N}} I_C(km+r)$. Since for every $m \in \mathbb{N}$ and $r \in B_m$ the value $\kappa(m,r) = Nm\phi(m,r) + r$ does not belong to C, we deduce that $\frac{1}{m} \sum_{r=0}^{m-1} \inf_{k \in \mathbb{N}} I_C(km+r) = 0$ for all m, and consequently $\underline{E}_{\text{res}}(C) = 0$.

On the other hand, $\underline{E}_\theta(C) = \lim_{n \to \infty} \inf_{k \geq 0} \frac{1}{n} \sum_{\ell=k}^{k+n-1} I_C(\ell)$. Since by construction any two elements in $\kappa(A)$ differ in at least N elements, we deduce that $\inf_{k \geq 0} \frac{1}{n} \sum_{\ell=k}^{k+n-1} I_C(\ell) \geq 1 - \frac{2}{N+1}$, and this for all $n \in \mathbb{N}$. This implies that $\underline{E}_\theta(C) \geq 1 - \frac{2}{N+1} > 0$. Hence, $\underline{E}_{\text{res}}$ is

---

[53]See De Cooman et al. [2006] for a proof.

strictly smaller than the smallest strongly shift-invariant natural extension $\underline{E}_\theta$, and therefore not strongly shift-invariant. ♦

8.2. **Strong T-invariance.** Now consider an arbitrary non-empty set $\mathscr{X}$. Also consider a transformation T of $\mathscr{X}$ and the Abelian monoid $\mathscr{T}_T = \{T^n : n \geq 0\}$ generated by T. We shall characterise the strongly T-invariant coherent lower previsions on $\mathscr{L}(\mathscr{X})$ using the Banach limits on $\mathscr{L}(\mathbb{N})$.

First of all, consider any coherent lower prevision $\underline{P}$ on $\mathscr{L}(\mathscr{X})$, and any gamble f on $\mathscr{X}$. Define the gamble $f_{\underline{P}}$ on $\mathbb{N}$ as

$$f_{\underline{P}}(n) := \underline{P}((T^t)^n f) = \underline{P}(f \circ T^n). \tag{16}$$

[This is indeed a gamble, as for all n we deduce from the coherence of $\underline{P}$ that $f_{\underline{P}}(n) = \underline{P}(f \circ T^n) \leq \sup[f \circ T^n] \leq \sup f$ and similarly $f_{\underline{P}}(n) \geq \inf f$.] On the one hand $(T^t f)_{\underline{P}}(n) = \underline{P}(T^t f \circ T^n)) = \underline{P}(T^t(f \circ T^n)) = f_{T\underline{P}}(n)$ and on the other hand $(T^t f)_{\underline{P}}(n) = \underline{P}(f \circ T^{n+1}) = f_{\underline{P}}(n+1) = f_{\underline{P}}(\theta n)$, so

$$(T^t f)_{\underline{P}} = f_{T\underline{P}} = \theta^t f_{\underline{P}}, \tag{17}$$

and this observation allows us to establish a link between the transformation T on $\mathscr{X}$ and the shift transformation θ on $\mathbb{N}$. This makes us think of the following trick, inspired by what Bhaskara Rao and Bhaskara Rao [1983, Section 2.1.3(9)] do for probability charges, rather than coherent lower previsions. Let L be any shift-invariant coherent prevision on $\mathscr{L}(\mathbb{N})$, or in other words, a Banach limit on $\mathscr{L}(\mathbb{N})$. Define the real-valued functional $\underline{P}_L$ on $\mathscr{L}(\mathscr{X})$ by $\underline{P}_L(f) := L(f_{\underline{P}})$. We show that this functional has very special properties.

Proposition 21. *Let L be a shift-invariant coherent prevision on $\mathscr{L}(\mathbb{N})$, let $\underline{P}$ be a coherent lower prevision on $\mathscr{L}(\mathscr{X})$, and let T be a transformation of $\mathscr{X}$. Then the following statements hold.*

1. *$\underline{P}_L$ is a weakly T-invariant coherent lower prevision on $\mathscr{L}(\mathscr{X})$ (with equality).*
2. *If $\underline{P}$ dominates a weakly T-invariant coherent lower prevision $\underline{Q}$ on $\mathscr{L}(\mathscr{X})$, then $\underline{P}_L$ dominates $\underline{Q}$.*
3. *If $\underline{P} = P$ is a coherent prevision, then P_L is a (strongly) T-invariant coherent prevision on $\mathscr{L}(\mathscr{X})$.*
4. *If $\underline{Q}$ is a weakly T-invariant coherent lower prevision on $\mathscr{L}(\mathscr{X})$, then the (strongly) T-invariant coherent prevision P_L dominates $\underline{Q}$ for any P in $\mathscr{M}(\underline{Q})$.*
5. *If $\underline{P} = P$ is a T-invariant coherent prevision, then $P_L = P$.*

Proof. We first prove the first statement. Consider gambles f and g on $\mathscr{X}$. Since $\inf f \leq f_{\underline{P}}$, it follows from the coherence of L that $\inf f \leq L(f_{\underline{P}}) = \underline{P}_L(f)$. Moreover, we have for any n in $\mathbb{N}$ that

$$(f+g)_{\underline{P}}(n) = \underline{P}((f+g) \circ T^n) = \underline{P}(f \circ T^n + g \circ T^n) \geq \underline{P}(f \circ T^n) + \underline{P}(g \circ T^n) = f_{\underline{P}}(n) + g_{\underline{P}}(n),$$

where the inequality follows from the coherence [super-additivity] of $\underline{P}$. Since L is coherent, we see that $\underline{P}_L(f+g) \geq L(f_{\underline{P}}) + L(g_{\underline{P}}) = \underline{P}_L(f) + \underline{P}_L(g)$. Finally, for any $\lambda \geq 0$, we have that $(\lambda f)_{\underline{P}}(n) = \underline{P}((\lambda f) \circ T^n) = \underline{P}(\lambda(f \circ T^n)) = \lambda \underline{P}(f \circ T^n) = \lambda f_{\underline{P}}(n)$, since $\underline{P}$ is coherent. Consequently $\underline{P}_L(\lambda f) = L(\lambda f_{\underline{P}}) = \lambda L(f_{\underline{P}}) = \lambda \underline{P}_L(f)$, since L is coherent. This proves that $\underline{P}_L$ is a coherent lower prevision on $\mathscr{L}(\mathscr{X})$ [because (P1)–(P3) are satisfied]. To show that it is weakly T-invariant, recall that $(T^t f)_{\underline{P}} = \theta^t f_{\underline{P}}$, whence

$$\underline{P}_L(T^t f) = L((T^t f)_{\underline{P}}) = L(\theta^t f_{\underline{P}}) = L(f_{\underline{P}}) = \underline{P}_L(f),$$

since L is shift-invariant.

To prove the second statement, assume that $\underline{P}$ dominates the weakly T-invariant coherent lower prevision $\underline{Q}$ on $\mathscr{L}(\mathscr{X})$. Then for any gamble f on $\mathscr{X}$, we see that

$$f_{\underline{P}}(n) = \underline{P}(f \circ T^n) \geq \underline{Q}(f \circ T^n) \geq \underline{Q}(f),$$

where the last inequality follows from the weak T-invariance of $\underline{Q}$. Consequently, since L is coherent, we get $\underline{P}_L(f) = L(f_{\underline{P}}) \geq \underline{Q}(f)$.

The third statement follows immediately from the first and the fact that P_L is a self-conjugate coherent lower prevision (and therefore a coherent prevision) because P and L are.

The fourth statement follows at once from the second and the third. The fifth is an immediate consequence of the definition of P_L. $\qquad\square$

We can use the results in this proposition to characterise all strongly T-invariant coherent lower previsions using Banach limits on $\mathscr{L}(\mathbb{N})$.

Theorem 22. *Let $\underline{P}$ be a weakly T-invariant coherent lower prevision defined on some T-invariant domain $\mathscr{K}$, that avoids sure loss. Then the set of all T-invariant coherent previsions on $\mathscr{L}(\mathscr{X})$ that dominate $\underline{P}$ on $\mathscr{K}$ is given by*

$$\{P_L \colon P \in \mathscr{M}(\underline{P}) \text{ and } L \in \mathbb{P}_\theta(\mathbb{N})\},$$

so the smallest strongly T-invariant coherent lower prevision $\underline{E}_{\underline{P},T}$ on $\mathscr{L}(\mathscr{X})$ that dominates $\underline{P}$, i.e., the strongly T-invariant natural extension of $\underline{P}$, is the lower envelope of this set, and also given by

$$\underline{E}_{\underline{P},T}(f) = \inf_{P \in \mathscr{M}(\underline{P})} \underline{E}_\theta(f_P) = \inf_{P \in \mathscr{M}(\underline{P})} \sup_{n \geq 1} \inf_{k \geq 0} \left[\frac{1}{n} \sum_{\ell=k}^{k+n-1} P((T^\ell)^t f) \right]$$

for any gamble f on $\mathscr{X}$. As a consequence, the set $\mathbb{P}_T(\mathscr{X})$ of all T-invariant coherent previsions on $\mathscr{L}(\mathscr{X})$ is given by

$$\mathbb{P}_T(\mathscr{X}) = \{P_L \colon P \in \mathbb{P} \text{ and } L \in \mathbb{P}_\theta(\mathbb{N})\}.$$

This tells us that all T-invariant coherent previsions can be constructed using Banach limits on $\mathscr{L}(\mathbb{N})$. The smallest strongly T-invariant coherent lower prevision $\underline{E}_T$ on $\mathscr{L}(\mathscr{X})$ is the lower envelope of this set, and also given by

$$\underline{E}_T(f) = \inf_{P \in \mathbb{P}(\mathscr{X})} \underline{E}_\theta(f_P) = \inf_{P \in \mathbb{P}(\mathscr{X})} \sup_{n \geq 1} \inf_{k \geq 0} \left[\frac{1}{n} \sum_{\ell=k}^{k+n-1} P((T^\ell)^t f) \right]$$

for any gamble f on $\mathscr{X}$.

Proof. First of all, a coherent prevision P on $\mathscr{L}(\mathscr{X})$ belongs to $\mathscr{M}(\underline{P})$, i.e., dominates $\underline{P}$ on its domain $\mathscr{K}$, if and only if P dominates the natural extension $\underline{E}_{\underline{P}}$ on all gambles. Moreover, $\underline{E}_{\underline{P}}$ is weakly T-invariant by Theorem 8. Now consider any $P \in \mathscr{M}(\underline{P})$. Use the above observations together with Proposition 21 [statements 3 and 4] to show that for any Banach limit L on $\mathscr{L}(\mathbb{N})$, P_L is a T-invariant coherent prevision that dominates $\underline{P}$. Conversely, if P is a T-invariant coherent prevision on $\mathscr{L}(\mathscr{X})$ that dominates $\underline{P}$ on $\mathscr{K}$, then by Proposition 21 [statement 5], $P = P_L$ for any Banach limit L on $\mathscr{L}(\mathbb{N})$. This shows that $\{P_L \colon P \in \mathscr{M}(\underline{P}), L \in \mathbb{P}_\theta(\mathbb{N})\}$ is indeed the set of T-invariant coherent previsions on $\mathscr{L}(\mathscr{X})$ that dominate $\underline{P}$ on $\mathscr{K}$. Consequently, $\underline{E}_{\underline{P},T}$ is the lower envelope of this set, whence for any gamble f on $\mathscr{X}$

$$\underline{E}_{\underline{P},T}(f) = \inf_{P \in \mathscr{M}(\underline{P})} \inf_{L \in \mathbb{P}_\theta(\mathbb{N})} P_L(f) = \inf_{P \in \mathscr{M}(\underline{P})} \inf_{L \in \mathbb{P}_\theta(\mathbb{N})} L(f_P)$$

and since $\underline{E}_\theta$ is the lower envelope of $\mathbb{P}_\theta(\mathbb{N})$,

$$= \inf_{P \in \mathscr{M}(\underline{P})} \underline{E}_\theta(f_P)$$

and using Eqs. (15) and (16),

$$= \inf_{P \in \mathscr{M}(\underline{P})} \sup_{n \geq 1} \inf_{k \geq 0} \left[\frac{1}{n} \sum_{\ell=k}^{k+n-1} P((T^\ell)^t f) \right].$$

The rest of the proof is now immediate.   □

8.3. **Generalised Banach limits.** The above results on monoids $\mathscr{T}_T$ generated by a single transformation T can be generalised towards more general monoids $\mathscr{T}$ of transformations of $\mathscr{X}$, such that the set $\mathscr{T}^*$ of convex mixtures of the lifted linear transformations in $\mathscr{T}^t$ is directed by the successor relation $\geqslant$ on $\mathscr{T}^*$. The following discussion establishes an interesting connection between strong invariance and the notion of a generalised Banach limit.

We can consider $\mathscr{T}^*$ as a monoid of transformations of itself, as follows: with any element T^* we associate a transformation of $\mathscr{T}^*$, also denoted by T^*, such that $T^*(S^*) := S^*T^* \in \mathscr{T}^*$, for any S^* in $\mathscr{T}^*$.[54] We can, in the usual fashion, lift T^* to a transformation $(T^*)^t$ on $\mathscr{L}(\mathscr{T}^*)$ by letting $(T^*)^t g = g \circ T^*$, or in other words

$$(T^*)^t g(S^*) = g(T^*(S^*)) = g(S^*T^*), \tag{18}$$

for any S^* in $\mathscr{T}^*$ and any gamble g on $\mathscr{T}^*$, i.e., $g \in \mathscr{L}(\mathscr{T}^*)$.

Now a *generalised Banach limit* [Schechter, 1997, Sections 12.33–12.38] on $\mathscr{L}(\mathscr{T}^*)$ is defined as any linear functional on $\mathscr{L}(\mathscr{T}^*)$ that dominates the limit inferior operator with respect to the directed set $\mathscr{T}^*$. Let us take a closer look at this limit inferior operator. It is defined by

$$\liminf_{\mathscr{T}^*} g = \liminf_{T^* \in \mathscr{T}^*} g(T^*) := \sup_{S^* \in \mathscr{T}^*} \inf_{T^* \geqslant S^*} g(T^*),$$

for any gamble g on $\mathscr{T}^*$. Now recall that $T^* \geqslant S^*$ if and only if there is some R^* in $\mathscr{T}^*$ such that $T^* = R^*S^*$, so we get, using Eq. (18), that

$$\liminf_{T^* \in \mathscr{T}^*} g(T^*) = \sup_{S^* \in \mathscr{T}^*} \inf_{R^* \in \mathscr{T}^*} g(R^*S^*) = \sup_{S^* \in \mathscr{T}^*} \inf_{R^* \in \mathscr{T}^*} (S^*)^t g(R^*) = \lim_{S^* \in \mathscr{T}^*} \underline{P}_{\mathscr{T}^*}((S^*)^t g),$$

where $\underline{P}_{\mathscr{T}^*}$ is the vacuous lower prevision on $\mathscr{L}(\mathscr{T}^*)$. If we look at Corollary 14 for the special case $\mathscr{X} = \mathscr{T}^*$ and the monoid of transformations $\mathscr{T}^*$, recall that we need to lift transformations in $\mathscr{T}^*$ before we can apply them to gambles, and that the lifted transformations of $\mathscr{T}^*$ already constitute a convex set[55], we easily get to the following conclusion.

Proposition 23. *The limit inferior operator on $\mathscr{L}(\mathscr{T}^*)$ is actually the point-wise smallest strongly $\mathscr{T}^*$-invariant coherent lower prevision on $\mathscr{L}(\mathscr{T}^*)$, and the generalised Banach limits on $\mathscr{L}(\mathscr{T}^*)$ are the $\mathscr{T}^*$-invariant coherent previsions on $\mathscr{L}(\mathscr{T}^*)$.*

---

[54]Usually, $T^*(S^*)$ is defined as T^*S^*, see for instance Walley [1991, Note 1 of Section 3.5.1]. But we have to take a different route here because the elements of $\mathscr{T}^*$ are convex mixtures of *lifted* transformations, and as we have seen, lifting reverses the order of application of transformations.

[55]In general, even if $\mathscr{T}^t$ is directed by the successor relation $\geqslant$, the limit inferior operator on $\mathscr{L}(\mathscr{T}^t)$ will not be strongly invariant. But convexification, or going from $\mathscr{T}^t$ to $\mathscr{T}^*$, makes the limit inferior strongly invariant. Observe in this respect that the limit inferior operator on $\mathscr{L}(\mathbb{N})$ is not strongly shift-invariant, but its 'convexified' counterpart $\underline{E}_\theta$ is.

We can now apply arguments similar to the ones in the previous section, for general monoids $\mathscr{T}$ of transformations of $\mathscr{X}$ such that $\mathscr{T}^*$ is directed. Consider any coherent lower prevision $\underline{P}$ on $\mathscr{L}(\mathscr{X})$ and any gamble f, and define the following gamble $f_{\underline{P}}$ on $\mathscr{T}^*$:

$$f_{\underline{P}}(S^*) := \underline{P}(S^*f)$$

for any S^* in $\mathscr{T}^*$, which generalises Eq. (16). Observe that, using Eq. (18),

$$(T^*f)_{\underline{P}}(S^*) = \underline{P}(S^*T^*f) = f_{\underline{P}}(S^*T^*) = (T^*)^t f_{\underline{P}}(S^*),$$

so

$$(T^*f)_{\underline{P}} = (T^*)^t f_{\underline{P}},$$

which generalises Eq. (17). If we consider any $\mathscr{T}^*$-invariant coherent prevision L on $\mathscr{L}(\mathscr{T}^*)$, or in other words a generalised Banach limit on $\mathscr{L}(\mathscr{T}^*)$, we can now define a new lower prevision $\underline{P}_L$ on $\mathscr{L}(\mathscr{X})$ by $\underline{P}_L(f) := L(f_{\underline{P}})$, and Proposition 21, as well as Theorem 22, can now easily be generalised from monoids of transformations with a single generator to arbitrary directed monoids. In particular, we find that

$$\underline{E}_{\underline{P},\mathscr{T}}(f) = \inf_{P \in \mathscr{M}(\underline{P})} \liminf_{T^* \in \mathscr{T}^*} P(T^*f) \text{ and } \underline{E}_{\mathscr{T}}(f) = \inf_{P \in \mathbb{P}(\mathscr{X})} \liminf_{T^* \in \mathscr{T}^*} P(T^*f)$$

for any gamble f on $\mathscr{X}$, where $\underline{P}$ is any weakly $\mathscr{T}$-invariant lower prevision that avoids sure loss.

9. PERMUTATION INVARIANCE ON FINITE SPACES

Assume now that $\mathscr{T}$ is a finite group $\mathscr{P}$ of permutations of $\mathscr{X}$. Then we have the following characterisation result for the weakly $\mathscr{P}$-invariant coherent lower previsions.

Theorem 24. *Let $\mathscr{P}$ be a finite group of permutations of $\mathscr{X}$. All weakly $\mathscr{P}$-invariant coherent lower previsions $\underline{Q}$ on $\mathscr{L}(\mathscr{X})$ have the form*

$$\underline{Q} = \frac{1}{|\mathscr{P}|} \sum_{\pi \in \mathscr{P}} \pi \underline{P}, \tag{19}$$

where $|\mathscr{P}|$ is the number of permutations in $\mathscr{P}$, and $\underline{P}$ is any coherent lower prevision on $\mathscr{L}(\mathscr{X})$.

Proof. Consider a coherent lower prevision $\underline{P}$ on $\mathscr{L}(\mathscr{X})$, and let $\underline{Q}$ be the corresponding lower prevision, given by Eq. (19). Then $\underline{Q}$ is coherent, as a convex mixture of coherent lower previsions $\pi\underline{P}$. Moreover, let ϖ be any element of $\mathscr{P}$, then

$$\varpi\underline{Q} = \frac{1}{|\mathscr{P}|} \sum_{\pi \in \mathscr{P}} (\varpi\pi)\underline{P}, = \frac{1}{|\mathscr{P}|} \sum_{\pi \in \varpi\mathscr{P}} \pi\underline{P},$$

where $\varpi\mathscr{P} = \{\varpi\pi : \pi \in \mathscr{P}\} = \mathscr{P}$, because $\mathscr{P}$ is a group of permutations. Consequently $\varpi\underline{Q} = \underline{Q}$, so $\underline{Q}$ is weakly $\mathscr{P}$-invariant.

Conversely, let $\underline{Q}$ be any weakly $\mathscr{P}$-invariant coherent lower prevision, then we recover $\underline{Q}$ on the left-hand side if we insert $\underline{Q}$ in the right-hand side of Eq. (19). So any weakly $\mathscr{P}$-invariant coherent lower prevision is indeed of the form (19). □

Next, we give an interesting representation result for the strongly $\mathscr{P}$-invariant coherent lower previsions, when in addition, $\mathscr{X}$ is a finite set.[56] As we shall see further on, this essentially simple result has many interesting consequences, amongst which a generalisation to coherent lower previsions of de Finetti's [1937] representation result for finite sequences of exchangeable random variables (see Section 9.2). Recall that $\mathscr{A}_\mathscr{P}$ is the set of all $\mathscr{P}$-invariant atoms of $\mathscr{X}$. For each A in $\mathscr{A}_\mathscr{P}$, define $P^u(\cdot|A)$ as the coherent prevision on $\mathscr{L}(\mathscr{X})$ all of whose probability mass is uniformly distributed over A, i.e., for all gambles f on $\mathscr{X}$:

$$P^u(f|A) = \frac{1}{|A|} \sum_{x \in A} f(x).$$

Finally, let $P^u(f|\mathscr{A}_\mathscr{P})$ denote the gamble on $\mathscr{A}_\mathscr{P}$ that assumes the value $P^u(f|\mathscr{A}_\mathscr{P})(A) := P^u(f|A)$ in any element A of $\mathscr{A}_\mathscr{P}$.

Theorem 25. *Let $\mathscr{P}$ be a group of permutations of the finite set $\mathscr{X}$. A coherent lower prevision on $\mathscr{L}(\mathscr{X})$ is strongly $\mathscr{P}$-invariant if and only if $\underline{P}(f) = \underline{P}_0(P^u(f|\mathscr{A}_\mathscr{P}))$ for all f in $\mathscr{L}(\mathscr{X})$, where $\underline{P}_0$ is an arbitrary coherent lower prevision on $\mathscr{L}(\mathscr{A}_\mathscr{P})$.*

Proof. We begin with the 'if' part. Let $\underline{P}_0$ be an arbitrary coherent lower prevision on $\mathscr{L}(\mathscr{A}_\mathscr{P})$, and suppose that $\underline{P} = \underline{P}_0(P^u(\cdot|\mathscr{A}_\mathscr{P}))$. Then it is easy to see that $\underline{P}$ is coherent. We show that $\underline{P}$ is strongly $\mathscr{P}$-invariant. Consider any gamble f on $\mathscr{X}$ and any $\pi \in \mathscr{P}$. Then for any A in $\mathscr{A}_\mathscr{P}$ and any gamble f on $\mathscr{X}$,

$$P^u(f - \pi^t f|A) = \frac{1}{|A|} \sum_{x \in A} [f(x) - f(\pi x)] = 0,$$

because $x \in A$ is equivalent to $\pi x \in A$. So we see that $\underline{P}(f - \pi^t f) = \underline{P}_0(0) = 0$, since $\underline{P}_0$ is coherent. In a similar way, we can prove that $\underline{P}(\pi^t f - f) = 0$, so $\underline{P}$ is indeed strongly $\mathscr{P}$-invariant.

To prove the 'only if' part, we first concentrate on the case of a $\mathscr{P}$-invariant coherent prevision P on $\mathscr{L}(\mathscr{X})$. Fix any gamble f on $\mathscr{X}$. Since P is a coherent prevision, we find that

$$f = \sum_{A \in \mathscr{A}_\mathscr{P}} f I_A \quad \text{and} \quad P(f) = \sum_{A \in \mathscr{A}_\mathscr{P}} P(f I_A) = \sum_{A \in \mathscr{A}_\mathscr{P}} P(f|A) P(A),$$

where we have used Bayes's rule to define $P(f|A) := P(f I_A)/P(A)$ if $P(A) > 0$ and $P(f|A)$ is arbitrary otherwise.

Now assume that P is $\mathscr{P}$-invariant. Fix any $\mathscr{P}$-invariant atom A in $\mathscr{A}_\mathscr{P}$ such that $P(A) > 0$ and let $\pi \in \mathscr{P}$. For any gamble f on $\mathscr{X}$, we see that $\pi^t(f I_A) = (\pi^t f) I_A$, since A is in particular π-invariant. Consequently

$$P(\pi^t f|A) = P((\pi^t f) I_A)/P(A) = P(\pi^t(f I_A))/P(A) = P(f I_A)/P(A) = P(f|A),$$

so $P(\cdot|A)$ is $\mathscr{P}$-invariant as well.[57] Now let for any y in the finite set A, $p(y|A) := P(\{y\}|A) \geq 0$, then on the one hand $\sum_{x \in A} p(x|A) = P(A|A) = 1$. On the other hand, it follows from the π-invariance of $P(\cdot|A)$ that $p(x|A) = p(\pi x|A)$ for any x in A. Since we know from Proposition 1 that $A = \{\pi x \colon \pi \in \mathscr{P}\}$, we see that $p(\cdot|A)$ is constant on A,

---

[56]We find the 'permutation symmetry' between Theorems 24 and 25 quite surprising: the former states that a weakly $\mathscr{P}$-invariant coherent lower prevision is a uniform prevision (or mixture) of coherent lower previsions, and the latter that a strongly $\mathscr{P}$-invariant coherent lower prevision is a coherent lower prevision of uniform previsions.

[57]This is an instance of a more general result, namely that coherent conditioning of a coherent lower prevision on an invariant event preserves both weak and strong invariance. A proof of this statement is not difficult, but outside the scope of this paper.

so $p(x|A) = 1/|A|$ for all x in A, and consequently $P(f|A) = P^u(f|A)$, whence $P(f) = \sum_{A \in \mathscr{A}_{\mathscr{P}}} P^u(f|A)P(A)$. So indeed there is a coherent prevision P_0 on $\mathscr{L}(\mathscr{A}_{\mathscr{P}})$, defined by $P_0(\{A\}) = P(A)$ for all $A \in \mathscr{A}_{\mathscr{P}}$, such that $P = P_0(P^u(\cdot|\mathscr{A}_{\mathscr{P}}))$.

Finally, let $\underline{P}$ be any strongly $\mathscr{P}$-invariant coherent lower prevision, so any $P \in \mathscr{M}(\underline{P})$ is $\mathscr{P}$-invariant and can therefore be written as $P = P_0(P^u(\cdot|\mathscr{A}_{\mathscr{P}}))$. If we let $\underline{P}_0$ be the (coherent) lower envelope of the set $\{P_0 : P \in \mathscr{M}(\underline{P})\}$, then since $\underline{P}$ is the lower envelope of $\mathscr{M}(\underline{P})$, we get immediately that $\underline{P} = \underline{P}_0(P^u(\cdot|\mathscr{A}_{\mathscr{P}}))$. $\qquad \square$

As an immediate corollary, we see that that the uniform coherent prevision P^u on $\mathscr{L}(\mathscr{X})$ is the only strongly $\mathscr{P}$-invariant coherent lower prevision on $\mathscr{L}(\mathscr{X})$ if and only if $\mathscr{X}$ is the only $\mathscr{P}$-invariant atom, i.e., if $\mathscr{A}_{\mathscr{P}} = \{\mathscr{X}\}$. This is for instance the case if $\mathscr{P}$ is the group of all permutations of $\mathscr{X}$, or more generally if $\mathscr{P}$ includes the cyclic group of permutations of $\mathscr{X}$. It should therefore come as no surprise that, since symmetry of beliefs is so often confused with beliefs of symmetry, the uniform distribution is so often (but wrongly so) considered to be a good model for complete ignorance.

Another immediate corollary of this result is that the smallest strongly $\mathscr{P}$-invariant coherent lower prevision on $\mathscr{L}(\mathscr{X})$ is given by $\underline{P}(f) = \inf_{A \in \mathscr{A}_{\mathscr{P}}} \frac{1}{|A|} \sum_{x \in A} f(x)$, which of course agrees with the uniform distribution when we let $\mathscr{P}$ be the group of all permutations.

These results do not extend to the case where we have transformations of $\mathscr{X}$ that are not permutations; as we have said before, as soon as we have two different constant transformations in the monoid $\mathscr{T}$, there are no strongly invariant belief models.

9.1. **A few simple examples.** We now apply the theorems above in a number of interesting and simple examples.

Example 5. Let $\mathscr{X} = \mathscr{X}_2 := \{1, 2\}$, then all coherent lower previsions on $\mathscr{L}(\mathscr{X}_2)$ are so-called *linear-vacuous mixtures*, i.e., convex combinations of a coherent (linear) prevision and the vacuous lower prevision, and therefore given by

$$\underline{P}(f) = \varepsilon[\alpha f(1) + (1 - \alpha)f(2)] + (1 - \varepsilon)\min\{f(1), f(2)\},$$

where $0 \le \alpha \le 1$ and $0 \le \varepsilon \le 1$. Let $\mathscr{P}_2$ be the set of all permutations of $\mathscr{X}_2$. Then the only strongly $\mathscr{P}_2$-invariant coherent lower prevision is the uniform coherent prevision

$$P_{\frac{1}{2}}(f) = \frac{1}{2}[f(1) + f(2)],$$

corresponding to $\alpha = \frac{1}{2}$ and $\varepsilon = 1$. The weakly $\mathscr{P}_2$-invariant coherent lower previsions are given by

$$\underline{P}(f) = \varepsilon P_{\frac{1}{2}}(f) + (1 - \varepsilon)\min\{f(1), f(2)\},$$

where $0 \le \varepsilon \le 1$, so they are all the convex mixtures of the uniform coherent prevision and the vacuous lower prevision. $\blacklozenge$

Example 6. Let $\mathscr{X} = \mathscr{X}_3 := \{1,2,3\}$, then all 2-monotone coherent lower previsions on $\mathscr{L}(\mathscr{X}_3)$ are given by[58]

$$\underline{P}(f) = m_1 f(1) + m_2 f(2) + m_3 f(3)$$
$$+ m_4 \min\{f(1), f(2)\} + m_5 \min\{f(2), f(3)\} + m_6 \min\{f(3), f(1)\}$$
$$+ m_7 \min\left\{\frac{f(1) + f(2)}{2}, \frac{f(2) + f(3)}{2}, \frac{f(3) + f(1)}{2}\right\}$$
$$+ m_8 \min\{f(1), f(2), f(3)\}.$$

where $0 \leq m_k \leq 1$ and $\sum_{k=1}^{8} m_k = 1$. Let $\mathscr{P}_3$ be the set of all permutations of $\mathscr{X}_3$. Then the only strongly $\mathscr{P}_3$-invariant coherent lower prevision is the uniform coherent prevision

$$P(f) = \frac{1}{3}[f(1) + f(2) + f(3)],$$

corresponding to $m_1 = m_2 = m_3 = \frac{1}{3}$ and $m_4 = m_5 = m_6 = m_7 = m_8 = 0$ [Observe that a coherent prevision is always 2-monotone.]. Weak $\mathscr{P}_3$- invariance, on the other hand, requires only that $m_1 = m_2 = m_3$ and $m_4 = m_5 = m_6$, so all the weakly $\mathscr{P}_3$-invariant and 2-monotone coherent lower previsions are given by

$$\underline{P}(f) = \frac{M_1}{3}[f(1) + f(2) + f(3)]$$
$$+ \frac{M_2}{3}[\min\{f(1), f(2)\} + \min\{f(2), f(3)\} + \min\{f(3), f(1)\}]$$
$$+ M_3 \min\left\{\frac{f(1) + f(2)}{2}, \frac{f(2) + f(3)}{2}, \frac{f(3) + f(1)}{2}\right\}$$
$$+ M_4 \min\{f(1), f(2), f(3)\}.$$

where $0 \leq M_k \leq 1$ and $M_1 + M_2 + M_3 + M_4 = 1$. The weakly $\mathscr{P}_3$-invariant and completely monotone coherent lower previsions (natural extensions of belief functions) correspond to the choice $M_3 = 0$. ♦

Example 7. Consider rolling a die for which there is evidence of symmetry between all even numbers, on the one hand, and between all odd numbers on the other. Let $\mathscr{X} = \mathscr{X}_6 := \{1,\ldots,6\}$ and let $\mathscr{P}_{eo}$ be the set of all permutations of $\mathscr{X}_6$ that map even numbers to even numbers and odd numbers to odd numbers. The $\mathscr{P}_{eo}$-invariant atoms are $\{1,3,5\}$ and $\{2,4,6\}$. By Theorem 25, the strongly $\mathscr{P}_{eo}$-invariant coherent previsions on $\mathscr{L}(\mathscr{X}_6)$, which are the precise belief models that are compatible with the subject's beliefs of symmetry, are given by

$$P(f) = \frac{\alpha}{3}[f(1) + f(3) + f(5)] + \frac{1 - \alpha}{3}[f(2) + f(4) + f(6)],$$

---

[58]An explicit proof of this statement is beyond the scope of this paper, but it runs along the following lines: (i) any coherent lower probability on the set of all events of a three-element space is 2-monotone [Walley, 1981, p. 58]; (ii) all 2-monotone coherent lower probabilities make up a convex set, and are convex mixtures of the extreme points of this set [Maaß, 2003, Chapter 2] (By the way, an argument similar to that in Maaß [2003, Chapter 2] shows that all strongly $\mathscr{T}$-invariant coherent lower previsions are (infinite) convex mixtures of the extreme strongly $\mathscr{T}$-invariant coherent lower previsions.); (iii) the 2-monotone coherent lower previsions on all gambles are natural extensions of the 2-monotone coherent lower previsions on all events [Walley, 1981, De Cooman et al., 2006, 2005b,a]; and (iv) natural extension to gambles of 2-monotone lower probabilities preserves convex mixtures.

where $0 \leq \alpha \leq 1$, and more generally, the strongly $\mathscr{P}_{\text{eo}}$-invariant coherent lower previsions on $\mathscr{L}(\mathscr{X}_6)$ are [apply Theorem 25 and use the results in Example 5]

$$\underline{P}(f) = \varepsilon \left[\frac{\alpha}{3}[f(1) + f(3) + f(5)] + \frac{1-\alpha}{3}[f(2) + f(4) + f(6)] \right]$$
$$+ (1-\varepsilon) \min \left\{ \frac{f(1) + f(3) + f(5)}{3}, \frac{f(2) + f(4) + f(6)}{3} \right\}$$

for $0 \leq \varepsilon \leq 1$ and $0 \leq \alpha \leq 1$. ♦

Example 8. Let us show that the point-wise smallest strongly invariant coherent lower prevision extension is not necessarily 2-monotone. Consider $\mathscr{X}_4 := \{1, 2, 3, 4\}$, and let π be the permutation of $\mathscr{X}_4$ defined by $\pi(1) = 2$, $\pi(2) = 1$, $\pi(3) = 4$ and $\pi(4) = 3$. Observe that π is its own inverse, so $\mathscr{T}_\pi = \{\text{id}_{\mathscr{X}_4}, \pi\}$ is a group. From Theorem 25 we infer that the point-wise smallest strongly π-invariant coherent lower prevision on all gambles is given by

$$\underline{E}_\pi(f) = \min \left\{ \frac{f(1) + f(2)}{2}, \frac{f(3) + f(4)}{2} \right\}.$$

Let us now consider the gambles f_1 and f_2 on $\mathscr{X}_4$, given by $f_1(1) = 0$, $f_1(2) = -1$, $f_1(3) = 1$, $f_1(4) = -1$ and $f_2(1) = -1$, $f_2(2) = -0.25$, $f_2(3) = -1.5$, $f_2(4) = 0$. Check that

$$\underline{E}_\pi(f_1 \wedge f_2) + \underline{E}_\pi(f_1 \vee f_2) = -1.25 - 0.125 = -1.375 < -0.5 - 0.75 = \underline{E}_\pi(f_1) + \underline{E}_\pi(f_2).$$

Hence, $\underline{E}_\pi$ is not 2-monotone. ♦

The following example shows that possibility measures are not very useful for modelling permutation invariance.

Example 9. Consider a possibility measure Π defined on all events of a finite space $\mathscr{X}$. Then there is a map $\lambda \colon \mathscr{X} \to \mathbb{R}^+$, called the *possibility distribution* of Π, such that $\lambda(x) := \Pi(\{x\})$ and moreover $\Pi(A) = \max_{x \in A} \lambda(x)$ for all non-empty events $A \subseteq \mathscr{X}$. We have mentioned before that Π is a coherent upper probability if and only if $\Pi(\mathscr{X}) = \max_{x \in \mathscr{X}} \lambda(x) = 1$. We shall assume this is the case. Now consider any group $\mathscr{P}$ of permutations of $\mathscr{X}$. Then clearly Π is weakly $\mathscr{P}$-invariant if and only if λ is constant on the $\mathscr{P}$-invariant atoms of $\mathscr{X}$. In particular, Π is weakly invariant with respect to all permutations if and only is λ is everywhere equal to one, so Π is the vacuous upper probability.

For strong $\mathscr{P}$-invariance, let $\overline{P}$ be any strongly $\mathscr{P}$-invariant coherent lower prevision whose domain contains at least all events. Let x be any element of $\mathscr{X}$, and let $[x]_\mathscr{P}$ be the $\mathscr{P}$-invariant atom that contains x. Then it follows from Theorem 25 that $\overline{P}(\{x\}) \leq 1/|[x]_\mathscr{P}|$. So for $\overline{P}$ to extend a possibility measure, it is necessary (but not sufficient) that there is at least one element z of $\mathscr{X}$ such that $\overline{P}(\{z\}) = 1$, implying that z should be left invariant by all the permutations in $\mathscr{P}$, or equivalently, $[z]_\mathscr{P} = \{z\}$. ♦

9.2. **Exchangeable lower previsions.** As another example, we now discuss the case of so-called exchangeable coherent lower previsions. Consider a non-empty finite set $\mathscr{X}_\kappa := \{1, \ldots, \kappa\}$ of categories, and N random variables $X_1, \ldots, X_N$ taking values in the same set $\mathscr{X}_\kappa$, where κ and N are natural numbers with $\kappa \geq 2$ and $N \geq 1$. The joint random variable $\mathbf{X} := (X_1, \ldots, X_N)$ assumes values in the set $\mathscr{X} := \mathscr{X}_\kappa^N$.[59] We want to model a subject's beliefs about the value that $\mathbf{X}$ assumes in $\mathscr{X}_\kappa^N$, and generally, we use a coherent lower prevision $\underline{P}$ on $\mathscr{L}(\mathscr{X}_\kappa^N)$ to represent such beliefs.

---

[59]This means that we assume these N random variables to be *logically independent*.

Now assume that our subject believes that all random variables X_k are generated by the same process at different times k, and that the properties of this process do not depend on the time k. So, the subject assesses that there is permutation symmetry between the different times k. How can such *beliefs of symmetry* be modelled?

With a permutation π of $\{1,\ldots,N\}$, we can associate (by the usual procedure of lifting) a permutation of $\mathscr{X} = \mathscr{X}_\kappa^N$, also denoted by π, that maps any $\mathbf{x} = (x_1,\ldots,x_N)$ in $\mathscr{X}_\kappa^N$ to $\pi\mathbf{x} := (x_{\pi(1)},\ldots,x_{\pi(N)})$. The belief models that are compatible with the subject's beliefs of symmetry, are therefore the coherent lower previsions on (subsets of) $\mathscr{L}(\mathscr{X}_\kappa^N)$ that are *strongly $\mathscr{P}_\kappa^N$-invariant*, where $\mathscr{P}_\kappa^N$ is the group of liftings to $\mathscr{X}_\kappa^N$ of all permutations of $\{1,\ldots,N\}$. Walley [1991, Chapter 9] calls such lower previsions *exchangeable*, as they generalise de Finetti's [1937] notion of exchangeable coherent previsions. We intend to characterise the exchangeable lower previsions using Theorem 25. This will lead us to a generalisation (Eq. (20)) of de Finetti's [1937] representation result for finite numbers of exchangeable random variables.

It should be mentioned here that we should, as always, clearly distinguish between 'beliefs of symmetry' and 'symmetry of beliefs'. The latter imposes much weaker requirements on coherent lower previsions, namely those of weak $\mathscr{P}_\kappa^N$-invariance, which is called *permutability* by Walley [1991, Chapter 9].[60] In particular, the permutation symmetry that goes along with ignorance can only be invoked to justify permutability, but not, of course, exchangeability. Observe in this respect that the vacuous lower prevision on $\mathscr{L}(\mathscr{X}_\kappa^N)$ is permutable, but not exchangeable. It is well-known (see for instance Zabell [1989a, 1992]), that Laplace's Rule of Succession can be obtained by updating a particular exchangeable coherent prevision, but it should be clear from the discussion in this paper that ignorance alone (the Principle of Insufficient Reason) cannot be invoked to justify using such an exchangeable prevision, as (with considerable hindsight) Laplace implicitly seems to have done (see for instance Howie [2002], Zabell [1989a, 1992]).

For any $\mathbf{x} = (x_1,\ldots,x_N)$ in $\mathscr{X}_\kappa^N$, the $\mathscr{P}_\kappa^N$-invariant atom $[\mathbf{x}]_{\mathscr{P}_\kappa^N}$ is the set of all permutations of (the components of) $\mathbf{x}$. If we define the set of possible *count vectors*

$$\mathscr{N}_\kappa^N = \left\{ (m_1,\ldots,m_\kappa)\colon m_k \in \mathbb{N}^+ \text{ and } \sum_{k=1}^{\kappa} m_k = N \right\}$$

and the *counting map* $\mathbf{T}\colon \mathscr{X}_\kappa^N \to \mathscr{N}_\kappa^N$ such that $\mathbf{T}(x_1,\ldots,x_N)$ is the κ-tuple, whose k-th component is given by

$$T_k(x_1,\ldots,x_N) = |\{\ell \in \mathscr{X}_\kappa\colon x_\ell = k\}|,$$

i.e., the number of components of x whose value is k, then the number of elements of the invariant atom $[\mathbf{x}]_{\mathscr{P}_\kappa^N}$ is precisely

$$\nu(\mathbf{T}(\mathbf{x})) := \binom{N}{T_1(\mathbf{x})\ldots T_\kappa(\mathbf{x})} = \frac{N!}{T_1(\mathbf{x})!\ldots T_\kappa(\mathbf{x})!}$$

and $\mathbf{T}$ is a bijection (one-to-one and onto) between $\mathscr{A}_{\mathscr{P}_\kappa^N}$ and $\mathscr{N}_\kappa^N$. An invariant atom is therefore completely identified by the count vector $\mathbf{T}(\mathbf{x})$ of any of its elements $\mathbf{x}$, and we shall henceforth denote the invariant atoms of $\mathscr{X}_\kappa^N$ by $[\mathbf{m}]$, where $\mathbf{m} = (m_1,\ldots,m_\kappa) \in \mathscr{N}_\kappa^N$, and $\mathbf{x} \in [\mathbf{m}]$ if and only if $\mathbf{T}(\mathbf{x}) = \mathbf{m}$.

---

[60]See Walley [1991, Chapter 9] for a much more detailed discussion of the difference between permutability and exchangeability.

The coherent prevision $P^u(\cdot|\mathbf{m})$ on $\mathscr{L}(\mathscr{X}_\kappa^N)$ whose probability mass is uniformly distributed over the invariant atom $[\mathbf{m}]$ is given by

$$P^u(f|\mathbf{m}) = \frac{1}{\nu(\mathbf{m})} \sum_{\mathbf{x} \in [\mathbf{m}]} f(\mathbf{x}).$$

Interestingly, this is the precise prevision that is associated with taking N a-select drawings without replacement from an urn with N balls, m_1 of which are of type 1, $\ldots$, and m_κ of which are of type κ. Theorem 25 now tells us that any exchangeable coherent lower prevision $\underline{P}$ on $\mathscr{L}(\mathscr{X}_\kappa^N)$ can be written as

$$\underline{P}(f) = \underline{P}_\kappa^N(P^u(f|\mathscr{N}_\kappa^N)), \tag{20}$$

where $\underline{P}_\kappa^N$ is some coherent lower prevision on $\mathscr{L}(\mathscr{N}_\kappa^N)$. This means that such *an exchangeable lower prevision can be associated with N a-select drawings from an urn with N balls of types $1, \ldots, \kappa$, whose composition $\mathbf{m}$ is unknown, but for which the available information about the unknown composition is modelled by a coherent lower prevision $\underline{P}_\kappa^N$.*

That exchangeable coherent previsions can be interpreted in terms of sampling without replacement from an urn with unknown composition, is actually well-known, and essentially goes back to de Finetti [1937]. Heath and Sudderth [1976] give a simple proof for random variables that may assume two values. But we believe our proof[61] for the more general case of exchangeable coherent *lower* previsions and random variables that may assume *more than two values*, is conceptually even simpler than Heath and Sudderth's proof, even though it is a special case of a much more general representation result (Theorem 25). The essence of the present proof in the special case of coherent previsions P is captured wonderfully well by Zabell's [1992, Section 3.1] succinct statement: "Thus P is exchangeable if and only if two sequences having the same frequency vector have the same probability."

Our subject's beliefs could, in addition, be symmetrical in the categories in $\mathscr{X}_\kappa = \{1, \ldots, \kappa\}$, for instance as a result of her ignorance about the process that generates the outcomes X_k at each time k. As we have seen, this will be typically represented by using a type of *weakly* invariant belief models, in this case with respect to permutations of the categories, rather than the times. Any permutation ϖ of $\mathscr{X}_\kappa$ induces a permutation of $\mathscr{X}_\kappa^N$, also denoted by ϖ, through

$$\varpi\mathbf{x} = \varpi(x_1, \ldots, x_N) := (\varpi(x_1), \ldots, \varpi(x_N)).$$

What happens if we require that $\underline{P}$, in addition to being exchangeable, should also be *weakly* invariant under all such permutations? It is not difficult to prove that

$$P^u(\varpi^{-1}f|\mathbf{m}) = P^u(f|\varpi\mathbf{m}),$$

where we let $\varpi\mathbf{m} = \varpi(m_1, \ldots, m_\kappa) := (m_{\varpi(1)}, \ldots, m_{\varpi(\kappa)})$ in the usual fashion. This implies that there is such weak invariance if and only if the coherent lower prevision $\underline{P}_\kappa^N$ on $\mathscr{L}(\mathscr{N}_\kappa^N)$ is *weakly* invariant with respect to all category permutations! In particular, this weak invariance is satisfied for the vacuous lower prevision on $\mathscr{L}(\mathscr{N}_\kappa^N)$. Another type of lower coherent prevision that exhibits such a combination of strong invariance for time permutations and weak invariance for category permutations, and which also has other very special and interesting properties, is constructed by taking lower envelopes of specific sets of Dirichlet-Multinomial distributions, leading to the so-called Imprecise Dirichlet-Multinomial Model (IDMM, see Walley and Bernard [1999]).

---

[61]Walley [1991, Chapter 9] also mentions this result for exchangeable coherent lower previsions. The essence of his argument is similar to what we do in the last paragraph of the proof of Theorem 25.

In the literature, however, it is sometimes required that a coherent precise prevision should be invariant with respect to the combined action of the permutations of times and categories. These are the so-called *partition exchangeable* previsions (see Zabell [1992] for an interesting discussion and historical overview). Of course, the generalisation of this notion to coherent lower previsions should be strongly invariant with respect to such combined permutations, and therefore be a lower envelope of partition exchangeable previsions. For such *partition exchangeable* lower previsions, Theorem 25 can be invoked to prove a representation result that is similar to that for coherent lower previsions that are only exchangeable. It should be clear that they correspond to exchangeable lower previsions for which the corresponding coherent lower prevision $\underline{P}_\kappa^N$ on $\mathscr{L}(\mathscr{N}_\kappa^N)$ is *strongly* rather than just weakly invariant with respect to all category permutations. Of course, any justification for such models should be based on beliefs that there is permutation symmetry in the categories behind the process that generates the outcomes X_k at different times k, and *cannot be justified by mere ignorance about this process.*

9.3. **Updating exchangeable lower previsions: predictive inference.**

Finally, let us discuss possible applications of the discussion in this paper to predictive inference. Assume that we have n^* random variables $X_1, \ldots X_{n^*}$, that may assume values in the set $\mathscr{X}_\kappa = \{1, \ldots, \kappa\}$. We assume that these random variables are assessed to be exchangeable, in the sense that any coherent lower prevision that describes the available information about the values that the joint random variable $\mathbf{X}^* = (X_1, \ldots, X_{n^*})$ assumes in $\mathscr{X}_\kappa^{n^*}$ should be exchangeable, i.e., strongly $\mathscr{P}_\kappa^{n^*}$-invariant. This requirement could be called *pre-data exchangeability*. So we know from the previous section that such a coherent lower prevision must be of the form $\underline{P} = \underline{P}_\kappa^{n^*}(P^u(\cdot|\mathscr{N}_\kappa^{n^*}))$, where $\underline{P}_\kappa^{n^*}$ is some coherent lower prevision on $\mathscr{L}(\mathscr{N}_\kappa^{n^*})$. We shall assume that $\underline{P}_\kappa^{n^*}$ is a lower envelope of a set of coherent previsions $\mathscr{M}_\kappa^{n^*}$ on $\mathscr{L}(\mathscr{N}_\kappa^{n^*})$.

Suppose we now observe the values $\mathbf{x} = (x_1, \ldots, x_n)$ of the first n random variables $\mathbf{X} = (X_1, \ldots, X_n)$, where $1 \leq n < n^*$. We ask ourselves how we should coherently update the belief model $\underline{P}$ to a new model $\underline{P}(\cdot|\mathbf{x})$ which describes our beliefs about the values of the remaining random variables $\mathbf{X}' = (X_{n+1}, \ldots, X_{n^*})$. This is, generally speaking, the problem of *predictive inference*. In order to make things as easy as possible, we shall assume that $\underline{P}(\{\mathbf{x}\}) > 0$, so our subject has some reason, prior to observing $\mathbf{x}$, to believe that this observation will actually occur, because she is willing to bet on its occurrence at non-trivial odds.

Let us denote by $n' = n^* - n$ the number of remaining random variables, then we know that $\mathbf{X}'$ assumes values in $\mathscr{X}_\kappa^{n'}$, and $\underline{P}(\cdot|\mathbf{x})$ will be a lower prevision on $\mathscr{L}(\mathscr{X}_\kappa^{n'})$.

We shall first look at the problem of updating the coherent prevision $P = Q(P^u(\cdot|\mathscr{N}_\kappa^{n^*}))$ for any coherent prevision Q in $\mathscr{M}_\kappa^{n^*}$. So consider any gamble g on $\mathscr{X}_\kappa^{n'}$. It follows from coherence requirements (Bayes's rule) that the updated coherent prevision $P(\cdot|\mathbf{x})$ is given by

$$P(g|\mathbf{x}) = \frac{P(gI_\mathbf{x})}{P(I_\mathbf{x})} = \frac{Q(P^u(gI_\mathbf{x}|\mathscr{N}_\kappa^{n^*}))}{Q(P^u(I_\mathbf{x}|\mathscr{N}_\kappa^{n^*}))}, \tag{21}$$

where $I_\mathbf{x}(\mathbf{x}^*) = 1$ if the first n components of the vector $\mathbf{x}^* \in \mathscr{X}_\kappa^{n^*}$ are given by the vector $\mathbf{x}$, and zero otherwise. Observe, by the way, that by assumption, $P(I_\mathbf{x}) \geq \underline{P}(I_\mathbf{x}) = \underline{P}(\{\mathbf{x}\}) > 0$.

Now for any $\mathbf{m}^*$ in $\mathscr{N}_\kappa^{n^*}$ we find that, with obvious notations,

$$P^u(gI_\mathbf{x}|\mathbf{m}^*) = \frac{1}{\nu(\mathbf{m}^*)} \sum_{\mathbf{T}'(\mathbf{x}')+\mathbf{m}=\mathbf{m}^*} g(\mathbf{x}') = \frac{\nu(\mathbf{m}^*-\mathbf{m})}{\nu(\mathbf{m}^*)} P^u(g|\mathbf{m}^*-\mathbf{m}) \tag{22}$$

where we let $\mathbf{m} = \mathbf{T}(\mathbf{x})$, and where $\mathbf{T}'$ maps samples $\mathbf{x}'$ in $\mathscr{X}_\kappa^{n'}$ to their corresponding count vectors $\mathbf{T}'(\mathbf{x}')$ in $\mathscr{N}_\kappa^{n'}$. Of course $\nu(\mathbf{m}^* - \mathbf{m})$ is non-zero only if $\mathbf{m}^* \geq \mathbf{m}$, or equivalently if $\mathbf{m}^* - \mathbf{m} \in \mathscr{N}_\kappa^{n'}$, or in other words if it is possible to select n balls of composition $\mathbf{m}$ without replacement from an urn with composition $\mathbf{m}^*$. In this expression, $P^u(\cdot|\mathbf{m}')$ stands for the coherent prevision on $\mathscr{L}(\mathscr{X}_\kappa^{n'})$ whose probability mass is uniformly distributed over the $\mathscr{P}_\kappa^{n'}$-invariant atom $[\mathbf{m}']$, for any $\mathbf{m}'$ in $\mathscr{N}_\kappa^{n'}$. Now for $g = 1$ we find that

$$P^u(I_\mathbf{x}|\mathbf{m}^*) = \frac{\nu(\mathbf{m}^* - \mathbf{m})}{\nu(\mathbf{m}^*)} = p(\mathbf{m}|\mathbf{m}^*) =: L_\mathbf{m}(\mathbf{m}^*) \tag{23}$$

is the probability of observing a sample of size n with composition $\mathbf{m}$ by sampling without replacement from an urn with composition $\mathbf{m}^*$. $L_\mathbf{m}$ is the corresponding likelihood function on $\mathscr{N}_\kappa^{n^*}$. We may as well consider $L_\mathbf{m}$ as a likelihood function on $\mathscr{N}_\kappa^{n'}$, and for any $\mathbf{m}'$ in $\mathscr{N}_\kappa^{n'}$ we let

$$L_\mathbf{m}(\mathbf{m}') := L_\mathbf{m}(\mathbf{m} + \mathbf{m}') = \frac{\nu(\mathbf{m}')}{\nu(\mathbf{m} + \mathbf{m}')}$$

be the probability that there remain n' balls of composition $\mathbf{m}'$ after drawing (without replacement) n balls of composition $\mathbf{m}$ from an urn with n^* balls. We may then rewrite Eq. (21), using Eqs. (22) and (23), as

$$P(g|\mathbf{x}) = \frac{Q(L_\mathbf{m} P^u(g|\mathscr{N}_\kappa^{n'}))}{Q(L_\mathbf{m})} = Q(P^u(g|\mathscr{N}_\kappa^{n'})|\mathbf{m}), \tag{24}$$

where $Q(L_\mathbf{m}) = P(I_\mathbf{x}) > 0$ by assumption, and $Q(\cdot|\mathbf{m})$ is the coherent prevision on $\mathscr{L}(\mathscr{N}_\kappa^{n'})$ defined by

$$Q(h|\mathbf{m}) := \frac{Q(L_\mathbf{m} h)}{Q(L_\mathbf{m})}, \tag{25}$$

for any gamble h on $\mathscr{N}_\kappa^{n'}$, i.e., $Q(\cdot|\mathbf{m})$ is the coherent prevision obtained after using Bayes's rule to update Q with the likelihood function $L_\mathbf{m}$. This means that *if Q is a belief model for the unknown composition of an urn with n^* balls, then $Q(\cdot|\mathbf{m})$ is the corresponding model for the unknown composition of the remaining n' balls in the urn, after n balls with composition $\mathbf{m}$ have been taken from it.*

Now if we have a coherent lower prevision $\underline{P}_\kappa^{n^*}$ on $\mathscr{L}(\mathscr{N}_\kappa^{n^*})$ that is a lower envelope of a set $\mathscr{M}_\kappa^{n^*}$ of coherent previsions Q, then coherence[62] tells us that the updated lower prevision $\underline{P}(\cdot|\mathbf{x})$ is precisely the lower envelope of the corresponding updated coherent previsions $P(\cdot|\mathbf{x})$, and consequently, using Eqs. (24) and (25), we find that

$$\underline{P}(g|\mathbf{x}) = \underline{P}_\kappa^{n^*}(P^u(g|\mathscr{N}_\kappa^{n'})|\mathbf{m}), \tag{26}$$

where $\underline{P}_\kappa^{n^*}(\cdot|\mathbf{m})$ is the coherent lower prevision on $\mathscr{L}(\mathscr{N}_\kappa^{n'})$ given by

$$\underline{P}_\kappa^{n^*}(h|\mathbf{m}) := \inf\left\{ \frac{Q(L_\mathbf{m} h)}{Q(L_\mathbf{m})} : Q \in \mathscr{M}_\kappa^{n^*} \right\} = \inf\left\{ Q(h|\mathbf{m}) : Q \in \mathscr{M}_\kappa^{n^*} \right\}, \tag{27}$$

for any gamble h on $\mathscr{N}_\kappa^{n'}$. In other words, $\underline{P}_\kappa^{n^*}(\cdot|\mathbf{m})$ is the coherent lower prevision obtained after using coherence (the so-called Generalised Bayes Rule) to update $\underline{P}_\kappa^{n^*}$ with the likelihood function $L_\mathbf{m}$. This means again that *if $\underline{P}_\kappa^{n^*}$ is a belief model for the unknown composition of an urn with n^* balls, then $\underline{P}_\kappa^{n^*}(\cdot|\mathbf{m})$ is the corresponding belief model for the unknown composition of the remaining n' balls in the urn, after n balls with composition $\mathbf{m}$ have been taken from it.*

[62]This follows from Walley's [1991, Section 6.5] Generalised Bayes Rule.

If we compare Eq. (26) with Eq. (20), we see that the updated belief model $\underline{P}(\cdot|\mathbf{x})$ is still strongly $\mathscr{P}_\kappa^{n'}$-invariant,[63] so there still is *post-data exchangeability* for the remaining random variables $\mathbf{X}' = (X_{n+1}, \ldots, X_{n^*})$. Moreover, by looking at Eq. (21) and Eqs. (26) and (27), we see that the updated (lower) previsions $P(\cdot|\mathbf{x})$ and $\underline{P}(\cdot|\mathbf{x})$ only depend on the observed sample $\mathbf{x}$ through the *likelihood function* $L_{\mathbf{T}(\mathbf{x})}$. This tells us that this type of predictive inference satisfies the so-called *likelihood principle*, and moreover that the count vector $\mathbf{m} = \mathbf{T}(\mathbf{x})$, or more generally the map $\mathbf{T}$ is a *sufficient statistic*.

10. CONCLUSIONS

We have tried to argue that there is a clear distinction between the symmetry of belief models, and models of beliefs of symmetry, and that both notions can be distinguished between when indecision is taken seriously, as is the case in Walley's [1991] behavioural theory of imprecise probabilities. Our present attempt to distinguish between these notions, and capture the distinction in a formal way, is inspired by Walley's [1991, Chapter 9] discussion of the difference between permutable and exchangeable lower previsions, and Pericchi and Walley's [1991] discussion of 'classes of reasonable priors' versus 'reasonable classes of priors'.

Indeed, there seems to be a difference of type between the two notions. The former (symmetry of models) is a property that belief models may have, and we may require, as a principle of rationality, or as a principle of 'faithful modelling', that if the available evidence is symmetrical, then our corresponding belief models should be symmetrical too. A case in point is that of complete ignorance, where the 'evidence' is completely symmetrical, and we may therefore require that corresponding belief model should be completely symmetrical too. This leads to the various principles discussed in Section 5, all of which seem to single out the vacuous belief model for representing complete ignorance, and which extend Walley's [1991, Section 5.5] treatment of this matter.

The latter notion (models of symmetry) is more properly related to a type of structural assessment: if a subject believes there is symmetry, how should she model that, and how should assessments of symmetry be combined with other assessments? We have tried to answer such questions in Sections 7, where we discuss the strongly invariant natural extension.

It is well-known that if we only use Bayesian, or precise, probability models, requiring invariance of the probability measures with respect to all types of symmetry in the evidence may be impossible; examples were given by Boole, Bertrand and Fisher (see Zabell [1989a] for discussion and references). This has led certain researchers to abandon requiring the above-mentioned 'faithfulness' of belief models, or to single out certain types of symmetry which are deemed to be better than others. We have tried to argue that this is unnecessary: the vacuous belief model has no such problems, and is symmetrical with respect to any transformation you care to name. And of course, our criticism of the Principle of Insufficient Reason is not new. Our ideas were heavily influenced by Walley's [1991] book on imprecise probabilities, whose Chapter 5 contains a wonderful overview of arguments against restricting ourselves to precise probability models. Zabell [1989b] also gives an excellent discussion of much older criticism, dating back to the middle of the 19th century. In particular, Ellis's [1844] *ex nihilo nihil* — you cannot make decisions or inferences based on complete ignorance — finds a nice confirmation in the fact that the vacuous belief model captures complete indecision, and that updating a vacuous belief model leads to a vacuous belief model [Walley, 1991, Section 6.6.1]. But what we have

---

[63]See also footnote 57.

tried to do here is provide a framework and mathematical apparatus that allows us to better understand and discuss the problems underlying the Principle of Insufficient Reason, and more general problems of dealing with any type of symmetry in belief models.

This study of symmetry in relation to belief models is far from being complete however, and our notions of weak and strong invariance may have to be refined, and perhaps even modified, as well as complemented by other notions of symmetry. It might for instance be of interest to study the notion of symmetry that captures the *insufficient reason to strictly prefer* that is briefly touched upon near the end of Section 4.1. Also, we may seem more certain than we actually are about the appropriateness (in terms of having a sound behavioural justification and interpretation) of our notions of weak and (especially) strong invariance for random variables that may assume an infinite number of values. This is the point where our intuition deserts us, and where a number of interesting questions and problems leave us speechless. To name but one such problem, brought to the fore by the discussion in Section 7: for certain types of monoids, it is completely irrational to impose strong invariance (because doing so makes us subject to a sure loss). We can understand why this is the case for the monoid of all transformations, even on a finite set (Theorem 6). But why, for instance, are there no (strongly) permutation invariant coherent (lower) previsions on the set of natural (and *a fortiori* real) numbers? Why are we (consequently) reduced to using (strong) shift or translation invariance of coherent (lower) previsions when we want to try and capture the idea of a uniform distribution on the set of natural (or real) numbers? And even then, why, as is hinted at in footnote 50, are there situations where updating a (strongly) shift-invariant coherent (lower) prevision produces a sure loss? Are there appropriately weakened versions of our strong invariance condition that avoid these problems?

REFERENCES

R. P. Agnew and A. P. Morse. Extensions of linear functionals, with applications to limits, integrals, measures and densities. *The Annals of Mathematics*, 39:20–30, 1938.

L. Alaoglu and G. Birkhoff. General ergodic theorems. *The Annals of Mathematics*, 41: 293–309, 1940.

K. P. S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges*. Academic Press, London, 1983.

M. M. Day. Ergodic theorems for Abelian semi-groups. *Transactions of the American Mathematical Society*, 51:399–412, 1942.

G. de Cooman. Integration and conditioning in numerical possibility theory. *Annals of Mathematics and Artificial Intelligence*, 32:87–123, 2001.

G. de Cooman. Belief models: an order-theoretic analysis. In G. de Cooman, T. L. Fine, and T. Seidenfeld, editors, *ISIPTA '01 – Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, pages 93–103. Shaker Publishing, Maastricht, 2000.

G. de Cooman. Belief models: an order-theoretic investigation. *Annals of Mathematics and Artificial Intelligence*, 45:5–34, 2005.

G. de Cooman and E. Miranda. Weak and strong laws of large numbers for coherent lower previsions. *Journal of Statistical Planning and Inference*, 2006. Submitted for publication.

G. de Cooman and M. C. M. Troffaes. Coherent lower previsions in systems modelling: products and aggregation rules. *Reliability Engineering and System Safety*, 85:113–134, 2004.

G. de Cooman and M. Zaffalon. Updating beliefs with incomplete observations. *Artificial Intelligence*, 159:75–125, 2004.

G. de Cooman, M. C. M. Troffaes, and E. Miranda. *n*-Monotone lower previsions and lower integrals. In F. G. Cozman, R. Nau, and T. Seidenfeld, editors, *ISIPTA 2005 – Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications*, pages 145–154. SIPTA, 2005a.

G. de Cooman, M. C. M. Troffaes, and E. Miranda. *n*-Monotone lower previsions. *Journal of Intelligent and Fuzzy Systems*, 16:253–263, 2005b.

G. de Cooman, M. C. M. Troffaes, and E. Miranda. *n*-Monotone exact functionals and their relation to lower and upper integrals. 2006. Submitted for publication.

B. de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937. English translation in Kyburg Jr. and Smokler [1964].

B. de Finetti. *Teoria delle Probabilità*. Einaudi, Turin, 1970.

B. de Finetti. *Theory of Probability*. John Wiley & Sons, Chichester, 1974–1975. English translation of de Finetti [1970], two volumes.

R. L. Ellis. On the foundations of the theory of probabilities. *Transactions of the Cambridge Philosophical Society*, 8:1–6, 1844. Reprinted in Walton [1863].

D. C. Heath and W. D. Sudderth. De Finetti's theorem on exchangeable variables. *The American Statistician*, 30:188–189, 1976.

D. Howie. *Interpreting Probability: Controversies and Developments in the Early Twentieth Century*. Cambridge Studies in Probability, Induction & Decision Theory. Cambridge University Press, Cambridge, UK, 2002.

J. B. Kadane and A. O'Hagan. Using finitely additive probability: uniform distributions on the natural numbers. *Journal of the American Statistical Association*, 90:636–631, 1995.

J. M. Keynes. *A Treatise on Probability*. Macmillan, London, 1921.

H. E. Kyburg Jr. and H. E. Smokler, editors. *Studies in Subjective Probability*. Wiley, New York, 1964. Second edition (with new material) 1980.

S. Maaß. *Exact functionals, functionals preserving linear inequalities, Lévy's metric*. PhD thesis, University of Bremen, 2003.

E. H. Moore and H. L. Smith. A general theory of limits. *American Journal of Mathematics*, 44:102–121, 1922.

S. Moral and N. Wilson. Revision rules for convex sets of probabilities. In G. Coletti, D. Dubois, and R. Scozzafava, editors, *Mathematical Models for Handling Partial Knowledge in Artificial Intelligence*, pages 113–128. Plenum Press, New York, 1995.

L. R. Pericchi and P. Walley. Robust Bayesian credible intervals and prior ignorance. *International Statistical Review*, 59:1–23, 1991.

E. Schechter. *Handbook of Analysis and Its Foundations*. Academic Press, San Diego, CA, 1997.

G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.

P. Walley. Coherent lower (and upper) probabilities. Statistics Research Report 22, University of Warwick, Coventry, 1981.

P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996a.

P. Walley. Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58:3–57, 1996b. With discussion.

P. Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125–148, 2000.

P. Walley and J.-M. Bernard. Imprecise probabilistic prediction for categorical data. Technical Report CAF-9901, Laboratoire Cognition et Activitées Finalisés, Université de Paris 8, January 1999.

W. Walton, editor. *The Mathematical and Other Writings of Rovert Leslie Ellis M.A.* Deighton and Bell, Cambridge, 1863.

S. L. Zabell. The Rule of Succession. *Erkenntnis*, 31:283–321, 1989a. Reprinted in Zabell [2005].

S. L. Zabell. R. A. Fisher on the history of inverse probability. *Statistical Science*, 4:247–256, 1989b. Reprinted in Zabell [2005].

S. L. Zabell. Predicting the unpredictable. *Synthese*, 90:205–232, 1992. Reprinted in Zabell [2005].

S. L. Zabell. *Symmetry and Its Discontents: Essays on the History of Inductive Probability*. Cambridge Studies in Probability, Induction, and Decision Theory. Cambridge University Press, Cambridge, UK, 2005.

L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

GHENT UNIVERSITY, SYSTeMS RESEARCH GROUP, TECHNOLOGIEPARK – ZWIJNAARDE 914, 9052 ZWIJNAARDE, BELGIUM

E-mail address: gert.decooman@ugent.be

REY JUAN CARLOS UNIVERSITY, DEP. OF STATISTICS AND OPERATIONS RESEARCH, C-TULIPÁN, S/N 28933, MÓSTOLES, SPAIN

E-mail address: enrique.miranda@urjc.es