6th International Symposium on Imprecise Probability: Theories and Applications, Durham, United Kingdom, 2009

Characterizing Factuality in Normal Form Sequential Decision Making

Nathan Huntley
Durham University
Department of Mathematical Sciences
Durham, UK
nathan.huntley@durham.ac.uk

Matthias C. M. Troffaes
Durham University
Department of Mathematical Sciences
Durham, UK
matthias.troffaes@gmail.com

Abstract

We prove necessary and sufficient conditions on choice functions for factuality to hold in normal form sequential decision problems. We find that factuality is sufficient for backward induction to work. However, choice must be induced by a total preorder for factuality to hold. Hence, many of the optimality criteria used in imprecise probability theory (such as interval dominance, maximality, and E-admissibility) are counterfactual under normal form decision making.

1 Introduction

Consider the two-stage decision problem depicted in Fig. 1. In the first stage, the subject chooses between either taking scones, or proceeding to the second stage. In the second stage, the subject chooses between either cake or ice cream. A normal form solution to this problem consists of the subject specifying all his admissible choices, at all stages, beforehand. One possible normal form solution is

$$\{\text{scones}, \text{no scones and then ice cream}\}.$$

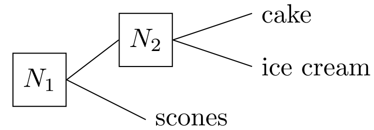

Figure 1: Two-stage problem.

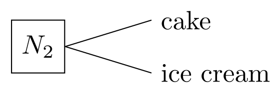

Figure 2: Second stage.

Imagine now that the subject already chose not to have scones. To resolve his choice between cake and ice cream, the subject can go back to the original problem that involved scones, and take the ice cream, but we might also imagine that he simply forgets about the scones and considers the simpler problem of choosing between cake and ice cream, as in Fig. 2.

If, faced with the simpler problem, the subject would now not state ice cream as his only admissible choice, we say that he is *counterfactual*: his choice between cake and ice cream depends on whether or not he had the choice of scones before. Perhaps, this seems an awkward property at first, but as we shall see, counterfactual choices are legion in many theories—a notable exception is maximizing expected utility.

So, when faced with a sequential decision problem, at any particular stage, one has two ways of looking at its normal form solution. Either, the problem can be thought of as part of a much larger problem (considering past choices one did not make and events that did not happen), or the problem can be thought of in its simplest form, not considering any past stages. Intuitively, a reasonable requirement is that the solution at any particular stage does not depend on the larger problem it is embedded in, i.e., that it is *factual*.

This papers studies necessary and sufficient conditions on choice functions for factuality in sequential decision problems when using normal form solutions, extending some results of Hammond [1] in his consequentialist theory. In doing so, factuality turns out to be sufficient for a backward induction scheme to work. We also find that choice must be induced by a total preorder for factuality to hold: for any choice function not induced by a total preorder, we can construct a counterfactual normal form example.

The relevance of this result for imprecise probability theory is that any criterion of optimality not induced by a total preorder (such as maximality, E-admissibility, and interval dominance) necessarily leads to counterfactuality. In other words, to satisfy factuality, one *must reject* either (i) the normal form as a means of solving decision problems, or (ii) any criterion that is not induced by a total preorder.

A total preorder, however, is not sufficient to imply factuality. Indeed, many total preorders that have been proposed for choice are still counterfactual. When precise probabilities are used, Hammond showed that expected utility is factual, as is well known, as are several related criteria [1, Sec. 9]. We are not aware of any non-trivial factual criteria that do not rely on probability and expected utility, although they may exist. The representation of all factual optimality criteria is still an open problem.

The paper is structured as follows: Section 2 explains decision trees and introduces notation. Section 3 provides a careful definition of normal and extensive form solutions, and introduces the concept of gambles to more easily work with normal form solutions. Section 4 introduces choice functions and their relationship with normal form solutions. Section 5 defines factuality and contains the principal results.

2 Decision Trees

2.1 Definition and Example

A decision tree [6] consists of a rooted tree of decision nodes, chance nodes, and reward leaves, growing from left to right. The left hand side corresponds to what happens first, and the right hand side to what happens last.

Consider Fig. 3. Decision nodes are depicted by squares, and chance nodes by circles. From each node, branches emerge. For decision nodes, each branch matches a decision; for chance nodes, each branch matches an event. For each chance node, the events that emerge form a partition of the possibility space: exactly one of the events must obtain. Each path in a decision tree corresponds to a particular sequence of decisions and events. The payoff resulting from each such sequence is put at the right end of the tree.

2.2 Notation

Decision trees can be seen as combinations of smaller decision trees: for instance, in the example, one could draw the subtree corresponding to d_S, and also draw the subtree corresponding to $d_{\overline{S}}$. The full decision tree then joins these two subtrees at a decision node.

Hence, we can represent a decision tree by its subtrees and the type of its root node. Let $T_1, \dots, T_n$ be decision trees and $E_1, \dots, E_n$ be a partition of the possibility space. If T is rooted at a decision node,

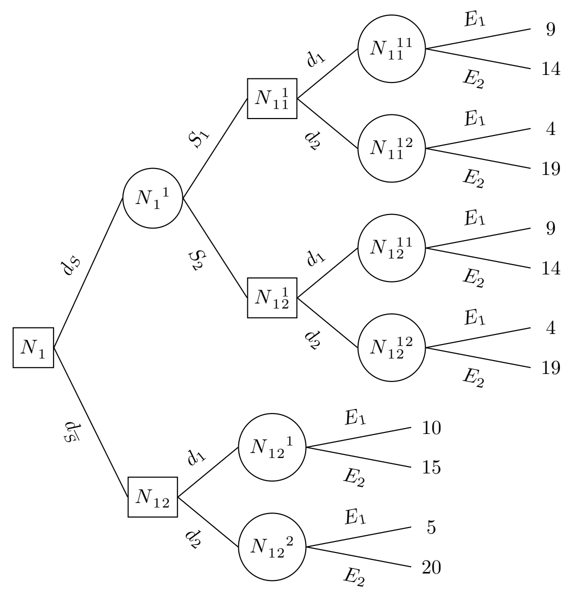

Figure 3: A decision tree.

we write $T = \bigsqcup_{i=1}^{n} T_i$, and at a chance node, we write $T = \bigodot_{i=1}^{n} E_i T_i$. For instance, for the tree of Fig. 3,

$$(S_1(T_1 \sqcup T_2) \odot S_2(T_1 \sqcup T_2)) \sqcup (U_1 \sqcup U_2) \text{ with}$$
$$T_1 = E_1 9 \odot E_2 14 \quad U_1 = E_1 10 \odot E_2 15$$
$$T_2 = E_1 4 \odot E_2 19 \quad U_2 = E_1 5 \odot E_2 20$$

Definition 1. *A subtree of a tree T obtained by removal of all non-descendants of a particular node N is called* the subtree of T at N *and is denoted by* $\mathrm{st}_N(T)$.

For any (sub)tree T, we summarize the events observed in the past as $\mathrm{ev}(T)$, which is the intersection of all the events on chance arcs that preceded T.

3 Solving Decision Trees

This paper deals with more general solutions of decision trees than are usually considered. Consequently, the standard definitions of extensive and normal forms, such as in Raiffa and Schlaifer [10], are insufficient for our purpose. Therefore, we first carefully define normal and extensive form solutions.

3.1 Extensive and Normal Form Solutions

An *extensive form solution* takes the decision tree and removes from each decision node some (possibly none), but not all, of the decision arcs. So, an extensive form solution is a subtree of the original decision tree, where at each decision node only a non-empty subset of arcs is retained. For instance, in the example, one of the extensive form solutions is: take $d_{\overline{S}}$, and then either take d_1 or d_2. An extensive form solution can be used as follows: the subject, upon reaching a decision node, chooses one of the arcs in the extensive form solution, and follows it. The subject only needs to decide which arc to follow at a decision node when reaching that node.

Following Raiffa and Schlaifer [10], Luce and Raiffa [7], and many others, another way to describe solutions to decision trees goes as follows. First, an extensive form solution with just one arc out of each decision node, is called a *normal form decision*. Hence, once a normal form decision is specified, a subject's decisions are uniquely determined in every eventuality. For instance, in the example, one of the normal form decisions is: take d_S, followed by d_1 if S_1 obtains, and d_2 if S_2 obtains. We denote the set of all normal form decisions for a decision tree T by $\mathrm{nfd}(T)$.

The interpretation of a normal form decision is that, upon reaching a decision node, the subject chooses the arc specified in the normal form decision. Compare this with a more general extensive form solution, in which the subject, upon reaching a decision node, chooses one of a subset of the available arcs. The difference between the two is that, for a normal form decision, the subject's choice at every decision node is uniquely determined from the beginning. In the extensive form, the particular arc to follow does not need to be determined unless the subject actually reaches the decision node in question.

A *normal form solution* of a decision tree T is then simply a subset of $\mathrm{nfd}(T)$. The interpretation of this subset is that the subject simply picks one of the normal form decisions of the normal form solution, and then acts accordingly.

Of course, an extensive form solution can always be transformed into a normal form solution by taking every possible normal form decision that is compatible with it. However, there are usually more normal form solutions than there are extensive form solutions.

3.2 Extensive and Normal Form Operators

An *extensive form operator* is a function which maps each decision tree to an extensive form solution of that decision tree. Note that some definitions in the literature, such as Raiffa and Schlaifer [10], define extensive form solutions through backward induction. Our definition does not specify the method by which decision arcs are removed. There need be no relationship between extensive forms and recursive methods.

An *normal form operator* is a function which maps each decision tree to a normal form solution of that decision tree. Again, note that the method by which this subset is determined is not part of our definition.

These operators usually (but do not need to) have the interpretation of describing optimal solutions.

An example of an extensive form operator is the classical backward induction method. Moving from right to left in the tree, decision arcs are deleted unless they give the maximum expected utility over all available arcs at that node. The principal feature of the method is that, once an arc has been deleted, it is ignored in all future calculations at nodes further to the left in the tree. The corresponding normal form operator finds the expected utility of each normal form decision and then returns the set that maximizes expected utility.

While it is well documented that these two classical operators always give equivalent solutions, this relationship can fail for other criteria. Extensive form operators that recursively apply a criterion may give a solution that differs from the normal form operator that applies the same criterion to the set of all normal form decisions. Examples can be found in Seidenfeld [11], Machina [8], and Jaffray [4], among others.

3.3 Gambles

In this paper we are primarily investigating normal form solutions. To express normal form decisions and solutions efficiently, we first introduce some definitions and notation. Let Ω be the *possibility space*: the set of all possible states of the world. We only consider finite possibility spaces. Elements of Ω are denoted by ω. Subsets of Ω are called *events*. The arcs emerging from chance nodes in a decision tree correspond to events.

Let $\mathcal{R}$ be a set of rewards. Often, rewards are measured in utiles, and hence $\mathcal{R} = \mathbb{R}$, but this assumption is not necessary for our results.

A *gamble* is a function $X\colon \Omega \to \mathcal{R}$; in other words, gambles are Ω–$\mathcal{R}$ functions. Gambles are interpreted as uncertain rewards: should $\omega \in \Omega$ be the true state of the world, the gamble X will yield the reward $X(\omega)$. Note that no probabilities over Ω are assumed at all.

3.4 Normal Form Gambles

Recall that a normal form decision prescribes the subject's actions, so once one has been chosen, the reward is determined entirely by the events that obtain. In other words, a normal form decision has a corresponding gamble, which we call a *normal form gamble*. The set of all normal form gambles associated with a decision tree T is denoted by $\mathrm{gamb}(T)$, so gamb is an operator on trees which yields the set of all gambles induced by normal form decisions of the tree.

	ω_1	ω_2	ω_3	ω_4
$E_1 9 \oplus E_2 14$	9	9	14	14
$S_1 (E_1 9 \oplus E_2 14)$ $\oplus S_2 (E_1 4 \oplus E_2 19)$	9	4	14	19

Table 1: Example of normal form gambles.

Let us explain how to find the gamble corresponding to a normal form decision, using Fig. 3 as an example. Instead of looking at the full tree, for simplicity let us first consider the subtree with root at N_{11}^1. The only two normal form decisions in this subtree are simply d_1 and d_2. The former gives reward 9 utiles if $\omega \in E_1$ and 14 utiles if $\omega \in E_2$, which corresponds to a gamble

$$E_1 9 \oplus E_2 14. \tag{1}$$

In the above expression, the $\oplus$ operator combines partial maps defined on disjoint domains (i.e. the constant partial map $E_1 9$ defined on E_1, and the constant partial map $E_2 14$ defined on E_2).

Now consider the subtree with root at ${N_1}^1$, and in particular the normal form decision 'd_1 if S_1 and d_2 if S_2'. This gives reward 9 if $\omega \in S_1 \cap E_1$, reward 14 if $\omega \in S_1 \cap E_2$, and so on. The corresponding gamble is

$$(S_1 \cap E_1)9 \oplus (S_1 \cap E_2)14 \oplus (S_2 \cap E_1)4 \oplus (S_2 \cap E_2)19,$$

or briefly, if we omit '$\cap$' and employ distributivity,

$$S_1 (E_1 9 \oplus E_2 14) \oplus S_2 (E_1 4 \oplus E_2 19),$$

where multiplication with an event is now understood to correspond to restriction, i.e., 9 is a constant map on Ω, $E_1 9$ is a constant map restricted to E_1, and $S_1(E_1 9)$ is obtained from $E_1 9$ by further restriction to $E_1 \cap S_1$. For illustration, we tabulate the values of some normal form gambles in Table 1, where $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $E_1 = \{\omega_1, \omega_2\}$, and $S_1 = \{\omega_1, \omega_3\}$.

Observe that the above gamble includes the gamble in Eq. (1) from N_{11}^1. Relationships between sets of normal form gambles for different subtrees allows a very convenient recursive definition of the gamb operator, given next. First, we extend $\oplus$ to sets of gambles:

Definition 2. *For any events $E_1, \ldots, E_n$ which form a partition, and any finite family of sets of gambles $\mathcal{X}_1, \ldots, \mathcal{X}_n$, we define the following set of gambles:*

$$\bigoplus_{i=1}^{n} E_i \mathcal{X}_i = \left\{ \bigoplus_{i=1}^{n} E_i X_i \colon X_i \in \mathcal{X}_i \right\}$$

Definition 3. *With any decision tree T, we associate a set of gambles* gamb(T)*, recursively defined through:*

- *If a tree T consists of only a leaf with reward $r \in \mathcal{R}$, then*

$$\mathrm{gamb}(T) = \{r\}. \tag{2a}$$

- *If a tree T has a chance node as root, that is, $T = \bigodot_{i=1}^{n} E_i T_i$, then*

$$\mathrm{gamb}\left(\bigodot_{i=1}^{n} E_i T_i \right) = \bigoplus_{i=1}^{n} E_i \, \mathrm{gamb}(T_i). \tag{2b}$$

- *If a tree T has a decision node as root, that is, if $T = \bigsqcup_{i=1}^{n} T_i$, then*

$$\mathrm{gamb}\left(\bigsqcup_{i=1}^{n} T_i \right) = \bigcup_{i=1}^{n} \mathrm{gamb}(T_i). \tag{2c}$$

Most decision problems can be modelled in more more than one way: there are usually multiple decision trees that model the same problem. This suggests the following definition (see for instance [8]):

Definition 4. *Two decision trees T_1 and T_2 are called* strategically equivalent *if* gamb(T_1) = gamb(T_2).

4 Normal Form Solutions for Decision Trees

4.1 Choice Functions and Optimality

A normal form solution of a decision tree T is a subset of the set nfd(T) of all its normal form decisions. Ideally one would like to identify a single normal form decision that the subject considers the best, but there is no reason to suppose that this is always possible. The subject might, however, still be able to identify some normal form decisions that he would never consider choosing, and eliminate these. This leaves a subset of normal form decisions that the subject would be willing to choose from. We say that the subject considers elements of this subset to be *optimal*.

For instance, in classical decision theory, each normal form decision induces a random real-valued gain, and assuming that all probabilities are fully specified, a normal form decision is considered optimal if its expected gain is maximized. As another example, consider the situation where the probabilities are not precisely known, but a set $\mathcal{M}$ of plausible probability distributions can be specified. Then the subject might consider as optimal any of those normal form decisions whose expected gain is maximal under at least one probability distribution in $\mathcal{M}$. In other situations one might use a different optimality criterion.

In these two examples, optimal decisions are determined by comparison of gambles. This is a common approach, and one we follow here, since we have seen that normal form decisions have corresponding gambles, and gambles are easier to work with. We therefore suppose that the subject has some way of determining an optimal subset of any set of gambles, conditional upon an event A (which corresponds to the $\mathrm{ev}(T)$ of the decision tree in question):

Definition 5. *A choice function* opt *is an operator that, for any non-empty event A, maps each non-empty finite set $\mathcal{X}$ of gambles to a non-empty subset of this set: $\emptyset \neq \mathrm{opt}(\mathcal{X}|A) \subseteq \mathcal{X}$.*

Note that common uses of choice functions in social choice theory, such as by Sen [12, p. 63, ll. 19–21] do not consider conditioning, and define choice functions for arbitrary sets of options (not for gambles only).

4.2 Normal Form Operator Induced by a Choice Function

We have seen that normal form decisions induce gambles, and have introduced choice functions, acting on sets of gambles, as a means to model optimality. Whence, we naturally arrive at a normal form operator $\mathrm{norm}_{\mathrm{opt}}$, simply by applying opt on the set of all gambles associated with the tree T and then finding the corresponding set of normal form decisions.

Definition 6. *Given any choice function* opt*, and any decision tree T with $\mathrm{ev}(T) \neq \emptyset$, we define*

$$\mathrm{norm}_{\mathrm{opt}}(T) = \{U \in \mathrm{nfd}(T)\colon \mathrm{gamb}(U) \subseteq \mathrm{opt}(\mathrm{gamb}(T)|\mathrm{ev}(T))\}.$$

Of course, since U is always a normal form decision, $\mathrm{gamb}(U)$ is always a singleton in this definition. In particular, the following equality holds,

$$\mathrm{gamb}(\mathrm{norm}_{\mathrm{opt}}(T)) = \mathrm{opt}(\mathrm{gamb}(T)|\mathrm{ev}(T)). \quad (3)$$

Note that, although $\mathrm{norm}_{\mathrm{opt}}$ is applied to trees, it really depends only on the set of normal form gambles associated with the tree. Hence, the operator $\mathrm{norm}_{\mathrm{opt}}$ will respect strategic equivalence:

Theorem 7. *If T_1 and T_2 are strategically equivalent, then $\mathrm{gamb}(\mathrm{norm}_{\mathrm{opt}}(T_1)) = \mathrm{gamb}(\mathrm{norm}_{\mathrm{opt}}(T_2))$ whenever $\mathrm{ev}(T_1) = \mathrm{ev}(T_2) \neq \emptyset$.*

If there are several strategically equivalent trees that are plausible representations of the same problem, the above theorem guarantees that our solution is independent of the particular representation we use.

When studying factuality, we consider $\mathrm{norm}_{\mathrm{opt}}$ for arbitrary subtrees of a given decision tree. To ensure that $\mathrm{norm}_{\mathrm{opt}}$ can be applied on each of such subtrees, the following condition is necessary:

Definition 8. *A decision tree T is called* consistent *if for every node N of T, $\mathrm{ev}(\mathrm{st}_N(T)) \neq \emptyset$.*

Clearly, if a decision tree T is consistent, then for any node N in T, $\mathrm{st}_N(T)$ is also consistent. We study only consistent decision trees because we consider $\mathrm{norm}_{\mathrm{opt}}(\mathrm{st}_N(T))$ for any node N in T, which is impossible when $\mathrm{ev}(\mathrm{st}_N(T)) = \emptyset$.

Usually, when constructing decision trees, one does not consider events which conflict with preceding events, hence consistency is satisfied. However, due to an oversight, some branch of a chance node might be connected to an event that cannot occur: such tree can always be made consistent by removing those nodes whose conditioning event is empty.

We sometimes need to know when a set of gambles can be represented by a consistent decision tree, conditional on some event. The following definition characterizes precisely those gambles:

Definition 9. *Let A be any non-empty event, and let $\mathcal{X}$ be a set of gambles. Then the following conditions are equivalent; if any (hence all) of them are satisfied, we say that $\mathcal{X}$ is A-*consistent*.*

(A) There is a consistent decision tree T with $\mathrm{ev}(T) = A$ and $\mathrm{gamb}(T) = \mathcal{X}$.

(B) For every $r \in \mathcal{R}$ and every $X \in \mathcal{X}$ such that $X^{-1}(r) \neq \emptyset$, it holds that $X^{-1}(r) \cap A \neq \emptyset$.

A gamble X is called A-consistent if $\{X\}$ is A-consistent.

5 Counterfactuals

We now give a discussion of issues arising from the use of operators, either normal form or extensive form, that use *counterfactual* reasoning, and find necessary and sufficient conditions on opt for $\mathrm{norm}_{\mathrm{opt}}$ to avoid counterfactuality. Counterfactual reasoning involves the consideration of events that did not occur or decisions that were not chosen. This is of interest because for many choice functions opt that have been suggested in the literature, $\mathrm{norm}_{\mathrm{opt}}$ is counterfactual.

5.1 Example and Definition

Counterfactuals are best illustrated by an example. Suppose we are applying an extensive form operator to the tree T in Fig. 3. This operator will delete some (possibly none) of the decision arcs at $N = N_{1\,1}^{\;1}$. If the choice of arcs to delete is influenced only by $\mathrm{st}_N(T)$ (that is, the operator would delete the same arcs at N regardless of the larger tree in which $\mathrm{st}_N(T)$ is embedded) then the operator is called *factual*. If the operator does not have this property (for instance, if the solution were to depend on the possible consequences of $d_{\overline{S}}$ or S_2), then it is called *counterfactual*.

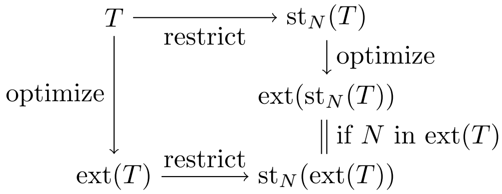

Figure 4: For a factual extensive form operator, optimization and restriction commute.

	A	$\overline{A}$
X	-1	-1
Y	-2	2
Z	0	0

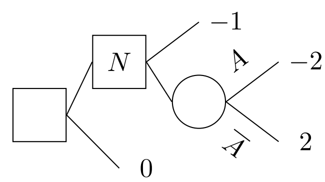

Figure 5: Decision tree for Example 12.

The definition of a counterfactual normal form operator requires the following extension to Definition 1.

Definition 10. *If $\mathcal{T}$ is a set of decision trees and N a node, then*

$$\mathrm{st}_N(\mathcal{T}) = \{\mathrm{st}_N(T) \colon T \in \mathcal{T} \text{ and } N \text{ in } T\}.$$

Definition 11. *An extensive form operator* ext *is called* factual *if for every consistent decision tree T and every node N such that N is in* $\mathrm{ext}(T)$*,*

$$\mathrm{st}_N(\mathrm{ext}(T)) = \mathrm{ext}(\mathrm{st}_N(T)),$$

otherwise, ext *is called* counterfactual*.*

An normal form operator norm *is called* factual *if for every consistent decision tree T and every node N such that N is in at least one element of* $\mathrm{norm}(T)$

$$\mathrm{st}_N(\mathrm{norm}(T)) = \mathrm{norm}(\mathrm{st}_N(T)),$$

otherwise, norm *is called* counterfactual*.*

In other words, for a factual operator, it does not matter whether we first restrict our attention to a subtree at a particular node N and then optimize this subtree, or first optimize, and only then look at the resulting subtree at a particular node N: roughly speaking, factuality means that optimization and restriction commute, as in Fig. 4 for an extensive form operator. For a counterfactual extensive form operator, $\mathrm{st}_N(\mathrm{ext}(T))$ can differ from $\mathrm{ext}(\mathrm{st}_N(T))$ for some decision trees T and nodes N in $\mathrm{ext}(T)$.

For example, the extensive form operator ext_P corresponding to the usual backward induction using expected utility is well known to be factual. Also, the usual normal form operator norm_P corresponding to maximizing expected utility over all normal form decisions is factual, because ext_P is equivalent to norm_P.

Before we examine factuality in more detail, we give an example of a counterfactual choice function.

Example 12. *Let T be the decision tree in Fig. 5, where X, Y, and Z are its normal form gambles. Under point-wise dominance, X and Y are incomparable, as are Y and Z. Hence,* $\mathrm{norm}(\mathrm{st}_N(T))$ *is $\{X, Y\}$ (where we conveniently identified normal form decisions with their normal form gambles). But* $\mathrm{norm}(T) = \mathrm{opt}(\{X, Y, Z\}) = \{Y, Z\}$ *as clearly Z dominates X. Restricting this solution to* $\mathrm{st}_N(T)$ *gives the normal form solution $\{Y\}$. Concluding,*

$$\{X, Y\} = \mathrm{norm}(\mathrm{st}_N(T)) \neq \mathrm{st}_N(\mathrm{norm}(T)) = \{Y\}$$

and therefore the normal form operator induced by point-wise dominance is counterfactual.

Even though point-wise dominance is counterfactual, it does satisfy $\mathrm{st}_N(\mathrm{norm}(T)) \subseteq \mathrm{norm}(\mathrm{st}_N(T))$, although this may not be true in general.

5.2 Necessary and Sufficient Conditions

In this section, we work extensively with normal form solutions, which are sets of trees. Therefore, it is convenient to extend gamb, $\odot$, and $\sqcup$, to sets of trees:

Definition 13. *For any set of decision trees $\mathcal{T}$,*

$$\mathrm{gamb}(\mathcal{T}) = \bigcup_{T \in \mathcal{T}} \mathrm{gamb}(T).$$

Definition 14. *For any sets of consistent decision trees $\mathcal{T}_1, \dots, \mathcal{T}_n$, and any partition $E_1, \dots, E_n$,*

$$\bigodot_{i=1}^{n} E_i \mathcal{T}_i = \left\{ \bigodot_{i=1}^{n} E_i T_i \colon T_i \in \mathcal{T}_i \right\}.$$

Definition 15. *For any sets of consistent decision trees $\mathcal{T}_1, \dots, \mathcal{T}_n$,*

$$\bigsqcup_{i=1}^{n} \mathcal{T}_i = \left\{ \bigsqcup_{i=1}^{n} T_i \colon T_i \in \mathcal{T}_i \right\}.$$

For sets of trees, the gamb operator satisfies:

$$\mathrm{gamb}\left(\bigodot_{i=1}^{n} E_i \mathcal{T}_i\right) = \bigoplus_{i=1}^{n} E_i \,\mathrm{gamb}(\mathcal{T}_i),$$

$$\mathrm{gamb}\left(\bigsqcup_{i=1}^{n} \mathcal{T}_i\right) = \bigcup_{i=1}^{n} \mathrm{gamb}(\mathcal{T}_i).$$

$$\mathrm{gamb}(T) = \mathrm{gamb}(\mathrm{nfd}(T)).$$

The following three properties turn out to be necessary and sufficient for factuality of normal form operators induced by a choice function.

Property 1 (Conditioning Property). *Let A be a non-empty event, and let $\mathcal{X}$ be a non-empty finite A-consistent set of gambles, with $\{X,Y\} \subseteq \mathcal{X}$ such that $AX = AY$. If $X \in \operatorname{opt}(\mathcal{X}|A)$, then $Y \in \operatorname{opt}(\mathcal{X}|A)$.*

Property 2 (Intersection property). *For any event $A \neq \emptyset$ and any non-empty finite A-consistent sets of gambles $\mathcal{X}$ and $\mathcal{Y}$ such that $\mathcal{Y} \subseteq \mathcal{X}$ and $\operatorname{opt}(\mathcal{X}|A) \cap \mathcal{Y} \neq \emptyset$, it holds that $\operatorname{opt}(\mathcal{Y}|A) = \operatorname{opt}(\mathcal{X}|A) \cap \mathcal{Y}$.*

For the next property, we use the following notation: if A is a non-trivial event (non-empty and not Ω), then $A\mathcal{X} \oplus \overline{A}Z = \{AX \oplus \overline{A}Z \colon X \in \mathcal{X}\}$.

Property 3 (Mixture property). *For any events A and B such that $A \cap B \neq \emptyset$ and $\overline{A} \cap B \neq \emptyset$, any $\overline{A} \cap B$-consistent gamble Z, and any non-empty finite $A \cap B$-consistent set of gambles $\mathcal{X}$,*

$$\operatorname{opt}(A\mathcal{X} \oplus \overline{A}Z|B) = A \operatorname{opt}(\mathcal{X}|A \cap B) \oplus \overline{A}Z.$$

Property 2 has a vast number of equivalent formulations, three of which we give next, yielding different interpretations to Property 2. These will be useful to discuss the implications of factuality later on.

Property 4 (Strong path independence). *For any non-empty event A and any non-empty finite A-consistent sets of gambles $\mathcal{X}_1, \ldots, \mathcal{X}_n$, there is a non-empty $\mathcal{I} \subseteq \{1, \ldots, n\}$ such that*

$$\operatorname{opt}\left(\bigcup_{i=1}^{n} \mathcal{X}_i \middle| A\right) = \bigcup_{i \in \mathcal{I}} \operatorname{opt}(\mathcal{X}_i|A)$$

Property 5 (Very strong path independence). *For any non-empty event A and any non-empty finite A-consistent sets of gambles $\mathcal{X}_1, \ldots, \mathcal{X}_n$,*

$$\operatorname{opt}\left(\bigcup_{i=1}^{n} \mathcal{X}_i \middle| A\right) = \bigcup_{\substack{i=1 \\ \mathcal{X}_i \cap \operatorname{opt}(\cup_{i=1}^{n} \mathcal{X}_i|A) \neq \emptyset}}^{n} \operatorname{opt}(\mathcal{X}_i|A)$$

Property 6 (Total preorder). *For every event $A \neq \emptyset$, there is a total preorder $\succeq_A$ on A-consistent gambles such that for every non-empty finite set of A-consistent gambles $\mathcal{X}$,*

$$\operatorname{opt}(\mathcal{X}|A) = \{X \in \mathcal{X} \colon (\forall Y \in \mathcal{X})(X \succeq_A Y)\}$$

Lemma 16. *Properties 2, 4, 5 and 6 are equivalent.*

To show that Properties 1, 2 and 3 are necessary and sufficient for factuality of $\operatorname{norm}_{\text{opt}}$, we require several lemmas (proofs are omitted due to space constraints).

We use this notation: for a decision tree T, $\operatorname{ch}(T)$ is the set of all child nodes of the root node of T.

Lemma 17. *Let norm be any normal form operator. Let T be a consistent decision tree. If,*

(i) for all nodes $K \in \operatorname{ch}(T)$ such that K is in at least one element of $\operatorname{norm}(T)$,

$$\operatorname{st}_K(\operatorname{norm}(T)) = \operatorname{norm}(\operatorname{st}_K(T)),$$

(ii) and, for all nodes $K \in \operatorname{ch}(T)$, and all nodes $L \in \operatorname{st}_K(T)$ such that L is in at least one element of $\operatorname{norm}(\operatorname{st}_K(T))$,

$$\operatorname{st}_L(\operatorname{norm}(\operatorname{st}_K(T))) = \operatorname{norm}(\operatorname{st}_L(\operatorname{st}_K(T))),$$

then, for all nodes N in T such that N is in at least one element of $\operatorname{norm}(T)$,

$$\operatorname{st}_N(\operatorname{norm}(T)) = \operatorname{norm}(\operatorname{st}_N(T)).$$

Lemma 18. *Let $A_1, \ldots, A_n$ be a finite partition of Ω, and let B be an event such that $A_i \cap B \neq \emptyset$ for all i. Let $\mathcal{X}_1, \ldots, \mathcal{X}_n$ be a finite family of non-empty finite sets of gambles, where $\mathcal{X}_i$ is $A_i \cap B$-consistent. If a choice function opt satisfies Properties 2 and 3, then*

$$\operatorname{opt}\left(\bigoplus_{i=1}^{n} A_i \mathcal{X}_i \middle| B\right) = \bigoplus_{i=1}^{n} A_i \operatorname{opt}(\mathcal{X}_i|A_i \cap B).$$

Lemma 19. *Consider a consistent decision tree T whose root is a decision node, so $T = \bigsqcup_{i=1}^{n} T_i$, and any choice function opt. For each tree T_i, let N_i be its root. Then, N_i is in at least one element of $\operatorname{norm}_{\text{opt}}(T)$ if and only if*

$$\operatorname{gamb}(T_i) \cap \operatorname{opt}(\operatorname{gamb}(T)|\operatorname{ev}(T)) \neq \emptyset.$$

Lemma 20. *For any consistent decision tree $T = \bigodot_{i=1}^{n} E_i T_i$, and any opt satisfying Property 1,*

$$\operatorname{gamb}(\operatorname{norm}_{\text{opt}}(T)) = \bigoplus_{i=1}^{n} E_i \operatorname{gamb}(\operatorname{norm}_{\text{opt}}(T_i))$$

implies

$$\operatorname{norm}_{\text{opt}}(T) = \bigodot_{i=1}^{n} E_i \operatorname{norm}_{\text{opt}}(T_i).$$

Lemma 21. *For any consistent decision tree $T = \bigsqcup_{i=1}^{n} T_i$ and any opt satisfying Property 2,*

$$\operatorname{gamb}(\operatorname{norm}_{\text{opt}}(T)) = \bigcup_{i \in \mathcal{I}} \operatorname{gamb}(\operatorname{norm}_{\text{opt}}(T_i)) \tag{4}$$

implies

$$\operatorname{norm}_{\text{opt}}(T) = \bigsqcup_{i \in \mathcal{I}} \operatorname{norm}_{\text{opt}}(T_i),$$

where $\mathcal{I} = \{i \colon \operatorname{gamb}(T_i) \cap \operatorname{opt}(\operatorname{gamb}(T)|\operatorname{ev}(T)) \neq \emptyset\}$.

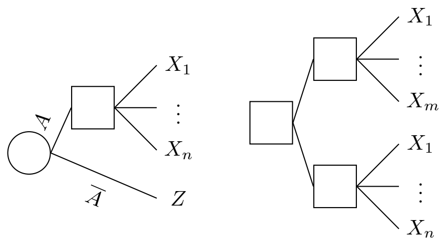

Figure 6: Decision trees for Theorem 22.

We are now ready to identify necessary and sufficient conditions for factuality.

Theorem 22. *A normal form operator* $\mathrm{norm}_{\mathrm{opt}}$ *is factual if and only if* opt *has Properties 1, 2 and 3.*

Proof. "only if". Omitting details, consider Fig. 6.

"if". We proceed by structural induction on all possible arguments of $\mathrm{norm}_{\mathrm{opt}}$, that is, on all consistent decision trees. In the base step, we prove the implication for trees consisting of only a single node. In the induction step, we prove that if the implication holds for the subtrees at every child of the root node, then the implication also holds for the whole tree.

First, if the decision tree T has only a single node, and hence, a reward at the root and no further children, then the condition for factuality is trivially satisfied.

Next, suppose that the consistent decision tree T has multiple nodes. Let $\{N_1, \dots, N_n\} = \mathrm{ch}(T)$, and let $T_i = \mathrm{st}_{N_i}(T)$. The induction hypothesis says that factuality is satisfied for all subtrees at every child of the root node, that is, for all T_i. More precisely, for all $i \in \{1, \dots, n\}$, and all nodes $L \in T_i$ such that L is in at least one element of $\mathrm{norm}_{\mathrm{opt}}(T_i)$

$$\mathrm{st}_L(\mathrm{norm}_{\mathrm{opt}}(T_i)) = \mathrm{norm}_{\mathrm{opt}}(\mathrm{st}_L(T_i)).$$

We must show that

$$\mathrm{st}_N(\mathrm{norm}_{\mathrm{opt}}(T)) = \mathrm{norm}_{\mathrm{opt}}(\mathrm{st}_N(T))$$

for all nodes N in T such that N is in at least one element of $\mathrm{norm}_{\mathrm{opt}}(T)$. By Lemma 17, and the induction hypothesis, it suffices to prove the above equality only for $N \in \mathrm{ch}(T)$, that is, it suffices to show that

$$\mathrm{st}_{N_i}(\mathrm{norm}_{\mathrm{opt}}(T)) = \mathrm{norm}_{\mathrm{opt}}(T_i) \tag{5}$$

for each $i \in \{1, \dots, n\}$ such that N_i is in at least one element of $\mathrm{norm}_{\mathrm{opt}}(T)$.

If T has a chance node as its root, that is, $T = \bigodot_{i=1}^n E_i T_i$, then all N_i are actually in every element of $\mathrm{norm}_{\mathrm{opt}}(T)$, so we must simply establish Eq. (5) for all $i \in \{1, \dots, n\}$. Equivalently, we must show that

$$\mathrm{norm}_{\mathrm{opt}}(T) = \bigodot_{i=1}^{n} E_i \, \mathrm{norm}_{\mathrm{opt}}(T_i) \tag{6}$$

Indeed, by Eq. (3),

$$\mathrm{gamb}(\mathrm{norm}_{\mathrm{opt}}(T)) = \mathrm{opt}(\mathrm{gamb}(T)|\mathrm{ev}(T))$$

and by the definition of the gamb operator, Eq. (2b) in particular,

$$= \mathrm{opt}\left(\bigoplus_{i=1}^{n} E_i \, \mathrm{gamb}(T_i) \middle| \mathrm{ev}(T)\right)$$

and so by Lemma 18,

$$= \bigoplus_{i=1}^{n} E_i \, \mathrm{opt}(\mathrm{gamb}(T_i)|\mathrm{ev}(T) \cap E_i)$$

so, since $\mathrm{ev}(T) \cap E_i = \mathrm{ev}(T_i)$, and again by Eq. (3),

$$= \bigoplus_{i=1}^{n} E_i \, \mathrm{gamb}(\mathrm{norm}_{\mathrm{opt}}(T_i))$$

Whence, Eq. (6) follows by Lemma 20.

Finally, assume that T has a decision node as its root, that is, $T = \bigsqcup_{i=1}^n T_i$. Let $\mathcal{I}$ be the subset of $\{1, \dots, n\}$ such that $i \in \mathcal{I}$ if and only if N_i is in at least one element of $\mathrm{norm}_{\mathrm{opt}}(T)$. We must establish Eq. (5) for all $i \in \mathcal{I}$. Equivalently, we must show that

$$\mathrm{norm}_{\mathrm{opt}}(T) = \bigsqcup_{i \in \mathcal{I}} \mathrm{norm}_{\mathrm{opt}}(T_i). \tag{7}$$

Indeed, by Eq. (3),

$$\mathrm{gamb}(\mathrm{norm}_{\mathrm{opt}}(T)) = \mathrm{opt}(\mathrm{gamb}(T)|\mathrm{ev}(T))$$

and by the definition of the gamb operator, Eq. (2c),

$$= \mathrm{opt}\left(\bigcup_{i=1}^{n} \mathrm{gamb}(T_i) \middle| \mathrm{ev}(T)\right)$$

and so by Property 5,

$$= \bigcup_{i \in \mathcal{I}^*} \mathrm{opt}(\mathrm{gamb}(T_i)|\mathrm{ev}(T)),$$

where $\mathcal{I}^* = \{i\colon\ \mathrm{gamb}(T_i) \cap \mathrm{opt}(\mathrm{gamb}(T)|\mathrm{ev}(T)) \neq \emptyset\}$, and so because $\mathrm{ev}(T) = \mathrm{ev}(T_i)$, and again by Eq. (3),

$$= \bigcup_{i \in \mathcal{I}^*} \mathrm{gamb}(\mathrm{norm}_{\mathrm{opt}}(T_i)).$$

Hence, the conditions of Lemma 21 are satisfied, and $\mathcal{I}^* = \mathcal{I}$ by Lemma 19, so Eq. (7) is established. $\square$

5.3 Backward Induction

A practical problem when solving decision trees using norm_{opt}, is that the set of normal form decisions of a tree T grows very fast with its size, and so $\text{gamb}(T)$ may have many elements. For this reason, elsewhere [3, 2], we have suggested the following backward induction method, which generalizes classical backward induction to arbitrary choice functions. To express this most conveniently, we first extend the norm_{opt} operator to act upon sets of decision trees.

Definition 23. *Given any set $\mathcal{T}$ of consistent decision trees, where $\text{ev}(T) = A$ for all $T \in \mathcal{T}$,*

$$\text{norm}_{\text{opt}}(\mathcal{T}) = \{U \in \text{nfd}(\mathcal{T}) : \text{gamb}(U) \subseteq \text{opt}(\text{gamb}(\mathcal{T})|A)\}.$$

Definition 24. *The normal form operator back_{opt} is defined for any consistent decision tree T through:*

- *If a tree T consists of only a leaf with reward $r \in \mathcal{R}$, then $\text{back}_{\text{opt}}(T) = \{T\}$.*
- *If a tree T has a chance node as root, that is, $T = \bigodot_{i=1}^{n} E_i T_i$, then*

$$\text{back}_{\text{opt}}(T) = \text{norm}_{\text{opt}}\left(\bigodot_{i=1}^{n} E_i \,\text{back}_{\text{opt}}(T_i)\right)$$

- *If a tree T has a decision node as root, that is, if $T = \bigsqcup_{i=1}^{n} T_i$, then*

$$\text{back}_{\text{opt}}(T) = \text{norm}_{\text{opt}}\left(\bigsqcup_{i=1}^{n} \text{back}_{\text{opt}}(T_i)\right).$$

If back_{opt} always yields the same normal form solution as norm_{opt}, we can use the former as an efficient way of calculating the latter. In [2] we show that the following four properties are necessary and sufficient for back_{opt} to coincide with norm_{opt}.

Property 7 (Backward conditioning property)**.** *Let A and B be events such that $A \cap B \neq \emptyset$ and $\overline{A} \cap B \neq \emptyset$, and let $\mathcal{X}$ be a non-empty finite $A \cap B$-consistent set of gambles, with $\{X, Y\} \subseteq \mathcal{X}$ such that $AX = AY$. Then $X \in \text{opt}(\mathcal{X}|A \cap B)$ implies $Y \in \text{opt}(\mathcal{X}|A \cap B)$ whenever there is a non-empty finite $\overline{A} \cap B$-consistent set of gambles $\mathcal{Z}$ such that, for at least one $Z \in \mathcal{Z}$,*

$$AX \oplus \overline{A}Z \in \text{opt}(A\mathcal{X} \oplus \overline{A}\mathcal{Z}|B).$$

Property 8 (Insensitivity of optimality to the omission of non-optimal elements)**.** *For any event $A \neq \emptyset$, and any non-empty finite A-consistent sets of gambles $\mathcal{X}$ and $\mathcal{Y}$,*

$$\text{opt}(\mathcal{X}|A) \subseteq \mathcal{Y} \subseteq \mathcal{X} \Rightarrow \text{opt}(\mathcal{Y}|A) = \text{opt}(\mathcal{X}|A).$$

Property 9 (Preservation of non-optimality under the addition of elements)**.** *For any event $A \neq \emptyset$, and any non-empty finite A-consistent sets of gambles $\mathcal{X}$ and $\mathcal{Y}$,*

$$\mathcal{Y} \subseteq \mathcal{X} \Rightarrow \text{opt}(\mathcal{Y}|A) \supseteq \text{opt}(\mathcal{X}|A) \cap \mathcal{Y}.$$

Property 10 (Backward mixture property)**.** *For any events A and B such that $B \cap A \neq \emptyset$ and $B \cap \overline{A} \neq \emptyset$, any $B \cap \overline{A}$-consistent gamble Z, and any non-empty finite $B \cap A$-consistent set of gambles $\mathcal{X}$,*

$$\text{opt}\left(A\mathcal{X} \oplus \overline{A}Z|B\right) \subseteq A\,\text{opt}(\mathcal{X}|A \cap B) \oplus \overline{A}Z.$$

Theorem 25 (Backward induction theorem [2])**.** *The following conditions are equivalent.*

(A) For any consistent decision tree T, it holds that $\text{back}_{\text{opt}}(T) = \text{norm}_{\text{opt}}(T)$.

(B) opt satisfies Properties 7, 8, 9, and 10.

Obviously, Property 1 implies Property 7, and Property 3 implies Property 10. Also,

Lemma 26. *Property 2 implies Properties 8 and 9.*

Hence, from Theorems 22 and 25, we conclude:

Corollary 27. *If norm_{opt} is factual, then $\text{norm}_{\text{opt}} = \text{back}_{\text{opt}}$.*

Factuality is, however, not necessary for backward induction. For example, it is easy to see that point-wise dominance satisfies Properties 1, 8, 9, and 10, but as we saw in Example 12, it is counterfactual.

Backward induction does imply a weaker version of factuality: $\text{st}_N(\text{norm}(T)) \subseteq \text{norm}(\text{st}_N(T))$.

5.4 Total Preordering

From Theorem 22 and Lemma 16, we have:

Corollary 28. *If norm_{opt} is factual then, for all $A \neq \emptyset$, $\text{opt}(\cdot|A)$ is induced by a total preorder.*

This constitutes a strong restriction on opt. Indeed, without consideration of factuality, a choice function that is not a total preorder may be desirable in some circumstances. When one has limited information about the relative likelihood of the events or the relative values of the rewards, one may wish to use a choice function that allows no preference between gambles, but does not consider them equivalent.

For example, if one is working with coherent lower previsions, one may consider the choice functions E-admissibility, maximality, and interval dominance, but none of these corresponds to a total preorder.

	Property						
	1	2	3	7	8	9	10
E-admissibility	✓		✓	✓	✓	✓	✓
Maximality	✓		✓	✓	✓	✓	✓
Γ-maximin	✓	✓		✓	✓	✓	
Interval Dominance	✓			✓	✓	✓	

Table 2: Properties of various choice functions.

Anyone wishing to use these choice functions to solve sequential decision problems must either abandon factuality or seek an alternative operator to norm_{opt}.

Those who prefer their choice functions to give a total preorder, on the other hand, can use factuality to justify this preference. Indeed, without consideration of factuality and sequential decisions, it is much harder to justify a total preorder than it is to justify simpler conditions such as Properties 8 and 9: see for instance Luce and Raiffa [7, pp. 288-289], where Axioms 6, 7, and $7''$ correspond to Properties 8, 9, and 2.

6 Conclusion

We defined factuality for extensive and normal forms. We found necessary and sufficient conditions for a choice function to induce a factual normal form operator. These turned out to be similar to, but stronger than, those for backward induction to work.

While many choice functions satisfy Property 1, Properties 2 and 3 are perhaps more restrictive than one would like. Is counterfactuality acceptable? We believe that factuality is a desirable property and one should think carefully before using a counterfactual operator. On the other hand, if one is attracted to the three properties for other reasons, then factuality gives them a strong justification.

Choice functions based on imprecise probability will typically violate at least one of Properties 2 and 3: Table 2 summarizes the properties satisfied by common choice functions. If one wishes to be factual in such cases, norm_{opt} cannot be used. Choice functions that induce factual extensive form operators are easier to find, particularly in the case of violations of Property 2 only: an example is sec_O in [11, p. 286]; also see Kikuti et al. [5]. Further investigation of factuality of extensive form operators, and in particular their relationships with backward induction and normal form operators, has been omitted due to lack of space.

Finally we mention that using counterfactuals is common in the field of causal inference [9]. This paper is quite different in character: we have not been concerned at all with causality and the use counterfactuals in causal inference. Instead, we have simply determined for what choice functions counterfactuals occur when solving decision trees.

Acknowledgements We thank the referees for useful feedback. EPSRC supports the first author.

References

[1] P. Hammond. Consequentialist foundations for expected utility. *Theory and Decision*, 25(1):25–78, Jul 1988.

[2] N. Huntley and M. C. M. Troffaes. Normal form backward induction for decision trees under arbitrary choice functions. Submitted.

[3] N. Huntley and M. C. M. Troffaes. An efficient normal form solution to decision trees with lower previsions. In *Soft Methods for Handling Variability and Imprecision*, Advances in Soft Computing, pages 419–426. Springer, Sep 2008.

[4] J. Jaffray. Rational decision making with imprecise probabilities. In *1st International Symposium on Imprecise Probabilities and Their Applications*, 1999.

[5] D. Kikuti, F. Cozman, and C. P. de Campos. Partially ordered preferences in decision trees: Computing strategies with imprecision in probabilities. In *IJCAI-05 Multidisciplinary Workshop on Advances in Preference Handling*, pages 118–123, 2005.

[6] D. V. Lindley. *Making Decisions.* Wiley, London, 2nd edition, 1985.

[7] R. D. Luce and H. Raiffa. *Games and Decisions: introduction and critical survery.* Wiley, 1957.

[8] M. J. Machina. Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature*, 27(1622-1688), 1989.

[9] S. L. Morgan and C. Winship. *Counterfactuals and causal inference.* Cambridge University Press, Cambridge, 2007.

[10] H. Raiffa and R. Schlaifer. *Applied Statistical Decision Theory.* Harvard University Press, 1961.

[11] T. Seidenfeld. Decision theory without 'independence' or without 'ordering': What is the difference? *Economics and Philosophy*, 4:267–290, 1988.

[12] A. K. Sen. Social choice theory: A re-examination. *Econometrica*, 45(1):53–89, 1977.