Proceedings of Machine Learning Research 147:342–351, 2021 ISIPTA 2021

The Sure Thing

Marco Zaffalon ZAFFALON@IDSIA.CH
IDSIA, Lugano, Switzerland

Enrique Miranda MIRANDAENRIQUE@UNIOVI.ES
Statistics and Operations Research, University of Oviedo, Spain

Abstract

If we prefer action a to b both under an event and under its complement, then we should just prefer a to b. This is Savage's *sure-thing principle*. In spite of its intuitive- and simple-looking nature, for which it gets almost immediate acceptance, the sure thing is not a logical principle. So where does it get its support from? In fact, the sure thing may actually fail. This is related to a variety of deep and foundational concepts in causality, decision theory, and probability, as well as to Simpsons' paradox and Blyth's game. In this paper we try to systematically clarify such a network of relations. Then we propose a general desirability theory for nonlinear utility scales. We use that to show that the sure thing is primitive to many of the previous concepts: In non-causal settings, the sure thing follows from considerations of temporal coherence and coincides with conglomerability; it can be understood as a rationality axiom to enable well-behaved conditioning in logic. In causal settings, it can be derived using only coherence and a causal independence condition.

Keywords: Probability, nonlinear utility scale, causality, Simpson's paradox, Blyth's game, temporal coherence, desirability, logic, coarsening, conglomerability.

1. Introduction

Not long ago we met a fascinating article by Pearl [15] on Savage's *sure thing principle* (STP, see [18]). Having worked in desirability for a long time, we kind-of took for granted that STP followed from the linearity property of coherent sets of gambles. But Pearl's take on the matter was different, in particular because his article connected STP with the quite recent research strand on causal inference.

The connection was somewhat upsetting. What has linearity in desirability to do with causality? Even worse, Pearl's causal argument gets naturally connected with Simpson's paradox [19]—as if Simpson's reversal is a failure of STP. But the paradox is compatible with probability, whence with desirability, and desirability yields STP via linearity. So how can STP fail in Simpson's? And are we talking of Simpson's or of the tightly related Blyth's game [4]? Moreover, actions are in the domain of decision theory, and we knew from our own research that desirability and decision theory are, to a large extent, one and the same thing [22, 24]. But then why should we need causality in the first place? Pearl also mentioned previous research by Aumann et al. [2] on the relation of STP with logic and incomplete observations.

We were confused to say the least.

This has motivated us to clarify the relations among such subjects. We introduce these subjects and their connections to STP in Section 3, after giving some preliminaries in Section 2. Sections 4 and 5 detail our analysis with special regard to Simpson's paradox and Blyth's game.

With a clearer mind, in Section 6 we move on to study what justifies the adoption of STP. To this end, we introduce a desirability theory generalised to nonlinear utility scales in Section 7. In Section 8 we discuss the foundations of STP in the non-causal setting; Section 9 explores them under causality. Our concluding views are in Section 10.

We are aware that not all the material has been fully clarified, now that this paper is finished, but it is organised more coherently. We see that as progress.[1]

2. Preliminaries

Given a space of possibilities Ω, a *gamble* $f:\Omega\to\mathbb{R}$ is a bounded real-valued function on Ω interpreted as an uncertain reward in a linear utility scale.

We denote by $\mathscr{L}(\Omega)$ the set of all the gambles on Ω and by $\mathscr{L}^+(\Omega) := \{f\in\mathscr{L}(\Omega) : 0\neq f\geq 0\}$ the subset of the *positive gambles*. We denote these sets also by $\mathscr{L}$ and $\mathscr{L}^+$, respectively, when there is no ambiguity about the space involved. Negative gambles are defined by $\mathscr{L}^- := -\mathscr{L}^+$.

Given a partition $\mathscr{B}$ of Ω, a gamble is said to be $\mathscr{B}$-*measurable* when it is constant on the elements of $\mathscr{B}$; we shall denote by $\mathscr{L}_{\mathscr{B}}$ the set of $\mathscr{B}$-measurable gambles.

A set of *desirable gambles* $\mathscr{D}\subseteq\mathscr{L}$ is called *coherent* if and only if the following conditions hold:

D1. $\mathscr{L}^+\subseteq\mathscr{D}$ [Accepting Partial Gains];

D2. $0\notin\mathscr{D}$ [Avoiding Status Quo];

D3. $f,g\in\mathscr{D}\Rightarrow f+g\in\mathscr{D}$ [Additivity];

D4. $f\in\mathscr{D},\lambda>0\Rightarrow\lambda f\in\mathscr{D}$ [Positive Homogeneity].

1. [20, p. ix].

For a deeper account of *desirability*, we refer to the work of Walley [20, Section 3.7] and Quaeghebeur [17].

In the following, events are denoted by capital letters such as $A,B,C \subseteq \Omega$; $\bar{A},\bar{B},\bar{C}$ denote their complements. We shall identify events with indicator functions, whence disjunctions $(A \cap B)$ will be represented by products (AB). Similarly, Bf is equal to f in B and zero elsewhere. It is interpreted as a conditional gamble: one that is called off if B does not occur. Symbol P will always denote a probability.

3. Some Main STP-Related Themes

Before delving into details, we should make clear that our notion of STP, informally given in the abstract, is not different from Savage's P2 [18, Section 2.7]. We should not read too much into this, though: e.g., desirability deals with zero-probability (or *null*) events differently from Savage's; this depends on questions other than P2, such as precision and Archimedeanity that Savage requires and we do not.

3.1. STP is Not the Logical STP

It is useful to clarify at the beginning that STP is not a logical principle, following some observations made by Aumann et al. [2, Section 7] (Savage said something similar by mentioning that STP is "extralogical").

We can be confused on this question since reasoning by cases in deductive logic may resemble STP quite closely. For instance, in propositional logic we have that for any four propositions A,B,B',C, it holds that

$$[((AB) \Rightarrow C) \wedge ((AB') \Rightarrow C)] \Rightarrow [(A(B \vee B')) \Rightarrow C]. \quad (1)$$

Note that by taking $B' := \bar{B}$, we get the form that resembles STP the most. And yet, the 'logical STP' given in Eq. (1), which is a tautology, is not STP.

One way to see this is the following: consider the experiment of tossing a die, with possibility space $A := \{1,2,3,4,5,6\}$. Let E be the event of getting number 1, while $B := \{1,2,3,4,5\}, B' := \{1,2,3,4,6\}$; then $P(E|B) = P(E|B') = 1/5$. Finally, let $C :=$"$P(E) = 1/5$". If we misinterpret probabilistic conditioning as a logical implication, Expression (1) leads us to conclude that $P(E) = 1/5$.[2]

In other words, as opposed to (1), actual STP can fail (see Section 5). Moreover, B and B' need not be a partition of unit, nor have they to be disjoint, for (1) to be true. This clashes with the actual STP, as detailed in Section 3.3.

3.2. STP and Probability

There is a tight relation between STP and probability theory once we look at it from the angle of desirability. Since conditioning a gamble f on an event B amounts to considering the product gamble Bf, we can re-write STP in this context as follows:

$$((\forall B \in \mathscr{B})\ Bf \in \mathscr{D} \cup \{0\}) \Rightarrow f \in \mathscr{D} \cup \{0\}. \quad (2)$$

This form of STP readily follows from the additivity axiom D3 of desirability in case the partition $\mathscr{B}$ is finite. Consider that the axioms D3 and D4 formalise the assumption that gambles express rewards in a linear utility scale. Whence we might be led to think that STP can just be regarded as an expression of a subject's linearity of the utility scale.

This view does not to stand up to a closer scrutiny. On the one hand, axiom D3 justifies far more than just (2): notably, it supports (2) for any collection of events B, whether they make up a partition or not. This is not the case of the actual STP (see Section 3.3). Stated differently, linearity of the utility scale appears to be supporting the logical STP rather than just STP. This is not too surprising as it is well known that D1–D4 can be regarded as axioms of a logic, thanks to some pioneering work done by de Cooman [5].

On the other hand, linearity of the utility scale is too restrictive in that it does not lead to STP when $\mathscr{B}$ is infinite. This is well known because in that case Eq. (2) coincides with the notion of 'conglomerability'. This is an important property to consider when conditioning a probabilistic model on an infinite partition $\mathscr{B}$ [7] (see also [3] for a review of such a concept in the literature). For this paper the question is important even only because the size of the partition seems to be irrelevant to the intuitive argument behind STP.

There are reasons to believe that conglomerability should be added as an additional axiom to D1–D4 [22, 24]. Could it be that STP is the underlying justification of conglomerability? But what is then the justification of STP?

3.3. STP Must Be Defined on Partitions

Aumann et al. [2] have shown that STP may fail if the possible observations B do not make up a partition $\mathscr{B}$ of Ω. Even though the authors do not point it out explicitly, the essence of the question they raise appears to be related to the assumption of 'coarsening at random' (CAR).

CAR is the (often implicit) assumption that enables us to deal with missing, and more generally coarsened, data, via conditioning. This has been discussed at length in a few papers published in the first decade of the century [6, 11, 21].

To represent the problem, besides the space of actual values Ω, one assumes that there is a space of possible observations Φ. These two spaces are related by a multi-valued map $\Gamma : \Omega \to \mathscr{P}(\Phi)$, which tells us what are the possible observations of actual value $\omega \in \Omega$.[3] For every $\phi \in \Phi$, let $\{\phi\}^* := \{\omega \in \Omega : \phi \in \Gamma(\omega)\}$, namely, all the elements of Ω that can lead to observation ϕ.

2. This type of observation was originally made by de Finetti [8]. We are grateful to a referee for pointing this out.

3. Not all maps Γ are compatible with CAR; see [21, Appendix A].

CAR holds if and only if

$$P(\phi|\omega) = P(\phi|\omega')$$

for all $\omega, \omega' \in \{\phi\}^*$ and all $\phi \in \Phi$ (see [21] for details).

At this point we can re-interpret Aumann's conclusion from the point of view of CAR quite easily. In fact, if the sets $\{\phi\}^*$ make up a partition of Ω, then there is exactly one observation in which each element $\omega \in \Omega$ can be mapped: whence, $P(\phi|\omega) = 1$ for all $\omega \in \{\phi\}^*$, and CAR follows.

This indicates that the observation by Aumann et al. [2] could be extended to the more general case where the observations satisfy CAR without making up a partition. We next prove that this is indeed the case. We focus on the case of precise probability to make things simpler.

Proposition 1 *Let $\Gamma : \Omega \to \mathscr{P}(\Phi)$ be a multivalued map representing a coarsening on Ω. Consider a gamble $f \in \mathscr{L}(\Omega)$, and assume that for every $\phi \in \Phi$, the gamble $f_\phi := \{\phi\}^* f$ is desirable. If CAR holds, then f is desirable.*

Proof CAR means that $P(\phi|\omega) =: \alpha_\phi$ is a constant for all $\omega \in \{\phi\}^* = \{\omega \in \Omega : \phi \in \Gamma(\omega)\}$. Since the gamble $f_\phi := \{\phi\}^* f$ is desirable for every $\phi \in \Phi$, it follows by D3 and D4 that so is the gamble $g := \sum_{\phi \in \Phi} \alpha_\phi f_\phi$. But actually for every $\omega \in \Omega$, it holds that

$$\begin{aligned} g(\omega) &= \sum_{\phi \in \Phi} \alpha_\phi \{\phi\}^*(\omega) f(\omega) = \sum_{\phi \in \Gamma(\omega)} \alpha_\phi f(\omega), \\ &= f(\omega) \sum_{\phi \in \Gamma(\omega)} P(\phi|\omega) = f(\omega), \end{aligned}$$

where the second equality follows because $\{\phi\}^*(\omega) = 1$ if and only if $\phi \in \Gamma(\omega)$, and the last is due to $P(\Gamma(\omega)|\omega) = 1$ for every ω by definition of the observational process. As a consequence $g = f$, whence f is desirable. ■

Otherwise in the paper we stick to the requirement that observations make up a partition.

3.4. STP and Causality

There is a final, and important, strand of research tightly related to STP, which is causal inference [13, 15].

Causality is a deep subject, but for this paper we just need to consider one of its aspects: namely, the fact that actions can change the 'world'; that is, actions implement a kind of model revision. We should bear in mind that this is different from probabilistic conditioning, which only focuses the attention on a specific part of the probabilistic model. To enforce the distinction, Pearl introduced the 'do' operator: $\mathrm{do}(X = x)$ denotes setting variable X to value x, as opposed to the case where X spontaneously takes value x, which is an event usually denoted simply by $(X = x)$.

Of course actions are in the domain of decision theory too. But traditional decision theory is not formulated in causal terms.

To see the source of this mismatch, we should consider that the axiomatisations of decision theory usually start by introducing the notion of a lottery. The lottery is made of many tickets for a number of prizes that we desire. In this context, actions are essentially interpreted as aimed at buying and selling lottery tickets. These transactions change the probability that we win a certain prize, but they do not change the probability that a certain ticket will eventually be drawn, i.e., they do not change the lottery itself. This setting, based on a lottery that is unaffected by our actions, suffices for the purpose of providing foundations for the notions of personal probability and utility [1]. But it entails no model revision. And this is the reason why causality does not enter the picture.

Yet, it is commonplace in real-world problems that our actions do change the underlying phenomenon. In these cases, Gibbard and Harper [10] have first observed, and Pearl [13, Theorem 6.1.1] has later shown, that for STP to work, relative to partition $\{B, \bar{B}\}$, the following causal independence condition is sufficient:[4]

$$P(B|\,\mathrm{do}(a)) = P(B|\,\mathrm{do}(b)). \tag{3}$$

Here a, b are any acts and B is the event under which, and also under its complement $\bar{B}$, we prefer a to b. In other words, our two possible actions should not structurally change the probability of event B.

3.5. STP and Simpson's Paradox

Simpson's paradox is a probabilistic phenomenon that has been puzzling scientists for a long time [19] (see also [14]). It occurs whenever there are three events A, B, C, so that:

$$P(C|A) > P(C|\bar{A}) \tag{4}$$
$$P(C|AB) < P(C|\bar{A}B) \tag{5}$$
$$P(C|A\bar{B}) < P(C|\bar{A}\bar{B}). \tag{6}$$

If C denotes recovery from a disease, B denotes gender, and A denotes placebo (versus treatment), for instance, then the paradox says that treatment is better than placebo both for men and women, but is worse overall (without knowing a person's gender). In other words, we witness a 'preference reversal' depending on what we decide to condition on.

So is the treatment good or bad? Stated differently, should we condition on gender or not? Pearl [13, Chapter 6] has argued at length that choosing the 'right' conditioning is a problem that cannot be solved within the boundaries of probability: we can instead solve it if we have a causal model of the domain.

The reason why Simpson's paradox is relevant to the discussion in this paper is that some people seem to regard it as a possible way to disprove STP.

4. It is well known that the problem can also be solved by assuming that acts are independent of states, as proposed by Jeffrey [12]. Jeffrey's condition is much stronger than Pearl's, though, see [15, p. 83].

The argument goes as follows: by Eq. (5), if B holds, we should prefer action a to b, where a is the act to bet on C when $\bar{A}$ is true, and b is the act to bet on C if A is true; a similar argument using Eq. (6) would mean that, if B does not hold, then we should prefer a to b. And then by STP we would deduce that a should be preferred to b, since this happens regardless whether B or $\bar{B}$ is observed, while Eq. (4) means that b should be preferred to a. The only way out seems to be that STP must break down.

Note however that Simpson's paradox is perfectly legitimate from a purely probabilistic point of view, as it is well known: Eqs. (4)–(6) do not contradict one another since they are compatible with a probabilistic model, as shown by numerous examples. And once the model exists, we can equivalently represent it as a coherent set of desirable gambles. And such a set needs to obey additivity (D3), from which STP follows.

So there appears to be a paradox about STP originated by Simpson's paradox. We detail the solution of such a paradox in Section 4.

4. Simpson Does Not Contradict STP

We shall argue here that the supposed failure of STP described in Section 3.5 is actually not a violation of STP. The flaw lies in an incorrect (not precise enough) translation of inequalities (4)–(6) into preferences. In what follows, we shall make things precise by equivalently translating those inequalities into a coherent set of desirable gambles $\mathscr{D}$.

The correspondence between coherent sets of desirable gambles and precise conditional probabilities is given by

$$\begin{aligned} P(A|B) &:= \sup\{\mu : B(A-\mu) \in \mathscr{D}\} \\ &= \inf\{\mu' : B(\mu' - A) \in \mathscr{D}\} \end{aligned} \tag{7}$$

for any pair of events A, B. (To simplify the exposition, we shall assume below that all probabilities are positive.)

We start by showing how to translate a generic probabilistic inequality into a desirability statement.

Lemma 2 *Consider events A,B,C,D. If $P(A|B) > P(C|D)$, then for any $\varepsilon \in (P(C|D), P(A|B))$ it holds that the gamble $AB - CD - \varepsilon(B - D)$ belongs to $\mathscr{D}$.*

Proof If $P(A|B) > P(C|D)$, then for any $\varepsilon \in (P(C|D), P(A|B))$ we deduce from Eq. (7) that $B(A-\varepsilon) \in \mathscr{D}$ and $D(\varepsilon - C) = -D(C-\varepsilon) \in \mathscr{D}$. Applying additivity (D3), we deduce that $B(A-\varepsilon) - D(C-\varepsilon) \in \mathscr{D}$, and this gamble is equal to $AB - CD - \varepsilon(B-D)$. ■

Lemma 2 has an important corollary:

Corollary 3 *Conditions (4)–(6) imply that for all $\varepsilon \in (P(C|\bar{A}), P(C|A))$, $\varepsilon_1 \in (P(C|AB), P(C|\bar{A}B))$, $\varepsilon_2 \in (P(C|A\bar{B}), P(C|\bar{A}\bar{B}))$, it holds that:*

$$AC - \bar{A}C - \varepsilon(A - \bar{A}) \in \mathscr{D} \tag{8}$$

$$\bar{A}BC - ABC - \varepsilon_1(\bar{A}B - AB) \in \mathscr{D} \tag{9}$$

$$\bar{A}\bar{B}C - A\bar{B}C - \varepsilon_2(\bar{A}\bar{B} - A\bar{B}) \in \mathscr{D}, \tag{10}$$

whence $(P(C|AB), P(C|\bar{A}B)), (P(C|\bar{A}), P(C|A))$ and $(P(C|A\bar{B}), P(C|\bar{A}\bar{B}))$ are pairwise disjoint.

Proof Eqs. (8)–(10) follow by a direct application of Lemma 2.

Next, let us assume without loss of generality that

$$P(C|AB) \leq P(C|A\bar{B}). \tag{11}$$

Then it must be $P(C|\bar{A}\bar{B}) > P(C|\bar{A}B)$: otherwise, $P(C|A)$, which is a convex combination of $P(C|AB)$ and $P(C|A\bar{B})$, cannot be greater than $P(C|\bar{A})$, which is a convex combination of $P(C|\bar{A}B)$ and $P(C|\bar{A}\bar{B})$. In other words, Eqs. (4)–(6) and (11) together imply that

$$P(C|AB) < P(C|\bar{A}B) < P(C|A\bar{B}) < P(C|\bar{A}\bar{B}).$$

On the other hand we have that $P(C|\bar{A}B) < P(C|\bar{A}) < P(C|\bar{A}\bar{B})$ and $P(C|AB) < P(C|A) < P(C|A\bar{B})$; that the inequalities are strict follows from the assumption that $P(AB), P(\bar{A},B), P(A,\bar{B}), P(\bar{A},\bar{B})$ are strictly positive. From this we deduce the chain of inequalities

$$\begin{aligned} P(C|AB) < P(C|\bar{A}B) < P(C|\bar{A}) < \\ P(C|A) < P(C|A\bar{B}) < P(C|\bar{A}\bar{B}), \end{aligned}$$

which completes the proof. ■

This leads to our main result for this section:

Theorem 4 *The three desirability assessments in Eqs. (8)–(10) do not imply a violation of coherence.*

Proof If we add Eqs. (9) and (10), we obtain that for all $\varepsilon_1 \in (P(C|AB), P(C|\bar{A}B))$ and all $\varepsilon_2 \in (P(C|A\bar{B}), P(C|\bar{A}\bar{B}))$,

$$\bar{A}C - AC - \varepsilon_1(\bar{A} - A) - (\varepsilon_2 - \varepsilon_1)(\bar{A}\bar{B} - A\bar{B}) \in \mathscr{D}.$$

If we add this with Eq. (8), we obtain that

$$-(\varepsilon - \varepsilon_1)(A - \bar{A}) - (\varepsilon_2 - \varepsilon_1)(\bar{A}\bar{B} - A\bar{B}) \in \mathscr{D} \tag{12}$$

for every $\varepsilon \in (P(C|\bar{A}), P(C|A))$, $\varepsilon_1 \in (P(C|AB), P(C|\bar{A}B))$, $\varepsilon_2 \in (P(C|A\bar{B}), P(C|\bar{A}\bar{B}))$. Note that, from Corollary 3, any such $\varepsilon, \varepsilon_1, \varepsilon_2$ satisfy $\varepsilon_1 < \varepsilon < \varepsilon_2$. We can rewrite (12) as

$$(\varepsilon_2 - \varepsilon)(A\bar{B} - \bar{A}\bar{B}) + (\varepsilon - \varepsilon_1)(\bar{A}B - AB) \in \mathscr{D}. \tag{13}$$

This does not contradict the coherence of $\mathscr{D}$, because such a gamble is always positive in $A\bar{B}$ and in $\bar{A}B$.

Let us establish now that no positive linear combination of gambles in (8)–(10) is smaller than or equal to zero. Fix $\varepsilon \in (P(C|\bar{A}), P(C|A))$, $\varepsilon_1 \in (P(C|AB), P(C|\bar{A}B))$, $\varepsilon_2 \in (P(C|A\bar{B}), P(C|\bar{A}\bar{B}))$ and denote the corresponding gambles $f_\varepsilon, f_{\varepsilon_1}, f_{\varepsilon_2}$. Assume ex-absurdo the existence of $\lambda_\varepsilon, \lambda_{\varepsilon_1}, \lambda_{\varepsilon_2} \geq 0$, not all of them zero, such that

$$f := \lambda_\varepsilon f_\varepsilon + \lambda_{\varepsilon_1} f_{\varepsilon_1} + \lambda_{\varepsilon_2} f_{\varepsilon_2} \leq 0.$$

If we look at the values taken by the combination above, we observe that, if $f \leq 0$, we should have

$$\lambda_\varepsilon(1-\varepsilon) = \lambda_{\varepsilon_1}(1-\varepsilon_1) = \lambda_{\varepsilon_2}(1-\varepsilon_2)$$

on the one hand and

$$\lambda_\varepsilon \varepsilon = \lambda_{\varepsilon_1}\varepsilon_1 = \lambda_{\varepsilon_2}\varepsilon_2$$

on the other. These two equations together imply that $\lambda_\varepsilon = \lambda_{\varepsilon_1} = \lambda_{\varepsilon_2}$, which means that we can reduce it to the case $\lambda_\varepsilon = \lambda_{\varepsilon_1} = \lambda_{\varepsilon_2} = 1$ considered before, for which we already obtained a contradiction with the incoherence of $\mathscr{D}$. ■

The point of the previous analysis is to detail why we cannot use Simpson's paradox to disprove STP. The key insight for this result is given by Corollary 3, as it shows that we are obliged to choose $\varepsilon < \varepsilon_1 < \varepsilon_2$. And without $\varepsilon = \varepsilon_1 = \varepsilon_2$, we cannot turn the gamble in (13) into zero.

To say it differently, Simpson's setting *seems* to contradict STP because there is an inversion of preferences; and we are intuitively tempted to believe that the inversion can be exploited as in Section 3.5 to yield a contradiction. But the formal translation of the inequalities into preferences (desirability) brings along constraints on the stakes $\varepsilon, \varepsilon_1, \varepsilon_2$ that make the reversal impossible to exploit to that end.

5. Blyth Does Contradict STP

We have seen that the bare use of Simpson's paradox does not lead to disprove STP, in spite of our intuition. But perhaps our intuition was not so wrong after all: in fact, all it takes to contradict STP is a twist on Simpson's that Blyth devised long ago [4].

Blyth considers a distribution of patients that satisfies Simpson's inequalities (4)–(6) and defines acts a and b, in the language of our medical example in Section 3.5, as follows:

a: Draw patients according to their distribution until you get one that received placebo (event A), and bet a dollar on the event C that the patient recovers.

b: Draw patients according to their distribution until you get one that received treatment (event $\bar{A}$) and bet a dollar on the event C that the patient recovers.

Event B denotes as before a person's gender.

Now the point is that whether the patient is male or female, we prefer b to a, according to (5)–(6). Then STP should imply that we prefer b to a without knowing a person's gender; but this contradicts (4). So STP fails here.

5.1. Causality Enters the Picture

The difference between Blyth's game and Simpson's paradox appears so thin that it is not easy to see what exactly makes STP fail in one case and hold in the other.

The solution to this further puzzle lies in Blyth's choice to sample patients until a condition is met, and only then place a bet. This makes acts a and b live in different, and incompatible, worlds: the first world is one where A is true; in the second world the complement event $\bar{A}$ is true instead.

The thin, but essential, difference with Simpson's is that these two worlds are not hypothetical, like in conditioning; those are worlds that actually occur, depending on whether we choose act a or act b. We will create a certain future, where A holds, by choosing a, and another future, where $\bar{A}$ holds, if we choose b. Those two future scenarios cannot both occur, by definition. Stated differently, both the first and the second worlds have probability one. For this reason, we cannot place a joint probability model over them.

The situation in Simpson's is different because it does not talk of future worlds. The most we do there is to reason conditionally on events occurring. But as it is well known, and as we detailed it in the past [22], conditioning is not about time; it only considers—at the same point in time—hypothetical scenarios. Each of these has a probability, and all of them together make up a joint (coherent) probabilistic model. So Simpson's does not fail STP because Simpson's paradox lives in a coherent world where additivity holds.

The last, but not least, fact about Blyth's is that the paradox is characterised by causality. Just because we 'do' act a or b: those are not observations, they are actions that we take (this is the very reason why we create two incompatible worlds).

And since causality enters the picture, we should ask ourselves, according to Pearl's condition (3), whether or not $P(B|\operatorname{do}(a)) = P(B|\operatorname{do}(b))$ as the validity of STP depends on such an equality. But the equality, in Blyth's game, translates into $P(B|A) = P(B|\bar{A})$ [15]. And it is well known that under the latter, Simpson's paradox is impossible to achieve. So $P(B|A) \neq P(B|\bar{A})$ and STP fails.

5.2. Probabilistic Equivalent Game

It is instructive to try to model Blyth's game probabilistically, as our (misguided) intuition might suggest. This requires representing Blyth's bets via conditioning:

- Betting one dollar on C in case A or $\bar{A}$ occur, with reference to Eq. (4): we represent these by the gambles $A(C-1)$ and $\bar{A}(C-1)$, respectively. Preferring the former to the latter translates into the gamble $h := A(C-1) - \bar{A}(C-1)$.
- Betting one dollar on C in case AB or $\bar{A}B$ occur, with reference to Eq. (5): gambles $AB(C-1)$ resp. $\bar{A}B(C-1)$. As for the preference we get the gamble: $h_B := \bar{A}B(C-1) - AB(C-1)$.

- Betting one dollar on C in case $A\bar{B}$ or $\bar{A}\bar{B}$ occur, with reference to Eq. (6): gamble $A\bar{B}(C-1)$ resp. $\bar{A}\bar{B}(C-1)$. Again, the preference is represented via $h_{\bar{B}} := \bar{A}\bar{B}(C-1) - A\bar{B}(C-1)$.

Now consider the expectation of h:

$$P(C-1|A)P(A) - P(C-1|\bar{A})P(\bar{A}). \tag{14}$$

Clearly we have a problem here. Because the expression that represents the difference in expectation of the two bets, with reference to inequality (4), is instead

$$P(C-1|A) - P(C-1|\bar{A}). \tag{15}$$

Note that the expressions (14) and (15) become the same if we set $P(A) := P(\bar{A}) := 1$. We cannot do this in (14) (Simpson's) because there we are representing the two hypothetical, coherent, worlds together. We instead have precisely that in (15) because there we are modelling the two actual, incompatible worlds, each one of those with an implicit probability equal to one.

But say that we still want to try to mimic Blyth's from Simpson's point of view. We can do this by rewriting gambles $h, h_B, h_{\bar{B}}$ as follows:

$$h' := A(C-1)/P(A) - \bar{A}(C-1)/P(\bar{A}).$$

$$h'_B := \bar{A}B(C-1)/P(\bar{A}B) - AB(C-1)/P(AB).$$

$$h'_{\bar{B}} := \bar{A}\bar{B}(C-1)/P(\bar{A}\bar{B}) - A\bar{B}(C-1)/P(A\bar{B}).$$

If we now take their expectations, those will coincide with the expectations of the three preferences that we obtain in Blyth's game. This means also that they are positive expectations, and hence that gambles $h', h'_B, h'_{\bar{B}}$ are desirable.

At this point, the complete equivalence with Blyth's game would follow if and only if $P(\bar{A}B) = P(AB)$ and $P(\bar{A}\bar{B}) = P(A\bar{B})$, and as a consequence $P(A) = P(\bar{A})$, because then also $h, h_B, h_{\bar{B}}$ would be desirable (use D4).

But these equalities imply in particular $P(B|A) = P(B|\bar{A})$, which is again Pearl's condition (3). Thus we eventually reach the same conclusion of Section 5.1: STP holds if and only if Blyth's does not.

6. Foundations of STP

At this point of the paper we have analysed STP from many different angles, but we are still left with the main open question: namely, why should STP be assumed as a principle of rationality? And in particular under which conditions, since we have seen that there are limits to its applicability—something that makes us wonder whether there may be others. To this end, it is convenient to consider the probabilistic and the causal STP separately.

We know that the probabilistic STP follows from additivity (D3), as discussed in Section 3.2. And since the same formalism of coherent sets of desirable gambles can be extended to decision-making à la Anscombe-Aumann [23, 24], we can obtain STP under utility conditions other than linearity, and hence in quite a general setting.

This approach is however unsatisfactory for two reasons: (i) as we said in Section 3.2, additivity does not seem to be a good candidate as the founding concept of probabilistic STP; and (ii) we would like to provide foundations for STP also in causal settings, which is not easy to do in the traditional decision-theoretic axiomatisations.

If we turn to causality, we can rely on Pearl's proof for STP in [13, Theorem 6.1.1]. The proof is nicely simple, given that the setting is deliberately based on probabilities over finite spaces. It shows that the essence of STP is total probability plus the causal independence condition (3).

Even though the result is nice and points to the very essence of the causal STP, we still regard Pearl's argument as partly satisfactory. The reason is that if we use probability to prove the causal STP, we are still implicitly relying on linearity of the utility scale to do so, that is, on axioms D3 and D4. For it is these axioms that are the 'core' of the axiomatisation of subjective probability; without them we would not have probability as we know it.

That the issue is linearity of the utility scale, seems to suggest, also in this case, moving towards those frameworks that deal with nonlinear utility: typically either Savage's [18] or Anscombe-Aumann's [1]. If we choose to pursue Savage's, the problem is that his approach assumes STP right from the start. So this route does not really seem to be accessible as it would lead to a circular argument. If we choose Anscombe-Aumann's, the problem (not too dissimilarly from Savage's) is that the underlying structure of their axiomatisation is linear; loosely speaking, they use compound lotteries so as to formalise nonlinear utility via a linear utility scale. This can be seen quite clearly from the fact that Anscombe-Aumann's framework is still based on axioms D3 and D4 (see in particular [23, Section 4.3]). Whence also this route is blocked.

Stated differently, we can as well extend Pearl's approach with considerations of expected utility, following traditional routes (as Pearl himself seems to suggest in the last paragraph of [15, Section 4]), but this will not really make the foundations of STP broader or deeper, given that we would implicitly use D3 and D4 again.

In order to overcome these limitations, we shall proceed as follows. In Section 7 we propose a generalised model of coherent sets of desirable gambles under nonlinear utility scales. This setting includes as a special case the usual formalisation of coherent sets of desirable gambles, and hence also Anscombe-Aumann's traditional, non-causal, formalisation of decision theory. In Section 8 we will show that under those generalised desirability conditions, STP can be derived for the non-causal setting. The causal setting is considered in Section 9.

7. Desirability with Nonlinear Utility Scales

We aim at defining coherent sets of desirable gambles under a very general notion of utility scales.

The utility scale is defined by the posi operator in the traditional definition of coherent sets of desirable gambles (see Section 2):

Definition 5 (Conic hull) *Given a set $\mathscr{D} \subseteq \mathscr{L}(\Omega)$, let*

$$\text{posi}(\mathscr{D}) := \left\{ \sum_{j=1}^{r} \lambda_j f_j : f_j \in \mathscr{D}, \lambda_j > 0, r \geq 1 \right\}$$

denote the conic hull *of the original set.*

When we regard the theory of desirability from a logical perspective, posi corresponds to the deductive closure; D1 to the tautologies and D2 to the status quo—which combined with the other axioms defines the contradictions, i.e., $\mathscr{L}^-$. (In the following we shall also use $\mathscr{L}_0^- := \mathscr{L}^- \cup \{0\}$.)

We proceed to generalise desirability by retaining the tautologies and the contradictions while replacing posi with the standard definition of a closure operator:

Definition 6 *(***Closure operator***) The map $\kappa : \mathscr{P}(\mathscr{L}) \to \mathscr{P}(\mathscr{L})$ is a* closure operator *if and only if for any two sets $\mathscr{G}, \mathscr{G}' \subseteq \mathscr{L}$ it satisfies:*

C1. $\mathscr{G} \subseteq \kappa(\mathscr{G})$ *[Extensiveness];*

C2. $\mathscr{G} \subseteq \mathscr{G}' \Rightarrow \kappa(\mathscr{G}) \subseteq \kappa(\mathscr{G}')$ *[Monotonicity];*

C3. $\kappa(\kappa(\mathscr{G})) = \kappa(\mathscr{G})$ *[Idempotency].*

Remark 7 *The closure operator* is *the utility scale in our approach. It seems worth stopping a moment to reflect on such a conceptual step. It tells us what is the 'hidden' reason-to-be of the utility scale: defining how desirable gambles relate to one another. More generally speaking, it appears to clarify the role of utility in the connection between probability and logic. For some reason, this does not seem to have been deepened very much so far, in our knowledge. In spite, for instance, of de Cooman [5] clearly pointing out long ago that the* posi *is a closure operator. There seems to be much to be gained in making this link explicit, both for logical theories at large and for desirability (and probability) in particular.*

Definition 8 *(***Natural extension***) Given a set $\mathscr{D}$ of desirable gambles, its* natural extension *$\mathscr{E}_\kappa(\mathscr{D})$ is the set of gambles given by $\mathscr{E}_\kappa(\mathscr{D}) := \kappa(\mathscr{D} \cup \mathscr{L}^+)$.*

The logical consistency of the assessments is characterised through the natural extension by the following:

Definition 9 *(***Avoiding partial loss***) A set $\mathscr{D}$ of desirable gambles is said to* avoid partial loss *if and only if $\mathscr{L}_0^- \cap \mathscr{E}_\kappa(\mathscr{D}) = \emptyset$.*

Their logical closure is instead defined by (κ-)coherence:

Definition 10 (Coherence relative to Q) *Say that $\mathscr{D}$ is* coherent relative to $Q \subseteq \mathscr{L}$ *if and only if $\mathscr{D}$ avoids partial loss and $Q \cap \mathscr{E}_\kappa(\mathscr{D}) \subseteq \mathscr{D}$ (and hence $Q \cap \mathscr{E}_\kappa(\mathscr{D}) = \mathscr{D}$).*

In the special case where $Q = \mathscr{L}$, $\mathscr{D}$ is simply said to be *coherent*, and it can be shown that:

Proposition 11 *$\mathscr{D}$ is coherent if and only if it satisfies the following conditions:*

K1. $\mathscr{L}^+ \subseteq \mathscr{D}$ *[Accepting Partial Gains];*

K2. $\mathscr{L}_0^- \cap \mathscr{D} = \emptyset$ *[Avoiding Partial Loss];*

K3. $\kappa(\mathscr{D}) = \mathscr{D}$ *[Deductive Closure].*

In the traditional theory κ is taken to be the posi. In that case, the avoiding partial loss condition reduces to $0 \notin \mathscr{E}_{\text{posi}}(\mathscr{D}) =: \mathscr{E}(\mathscr{D})$ and coherence simplifies to D1–D4.

Definition 12 (Conditioning) *Consider a coherent set of desirable gambles $\mathscr{D}$ on $\mathscr{L}$ and let B be a nonempty subset of Ω. Let $\mathscr{L}|B := \{Bf : f \in \mathscr{L}\}$ be the gambles that equal zero outside B. The set $\mathscr{D}$ conditional on B is defined as*

$$\mathscr{D}|B := \mathscr{D} \cap \mathscr{L}|B = \{f \in \mathscr{D} : f = Bf\}.$$

The conditional set is coherent relative to $\mathscr{L}|B$.

Conditional sets of gambles are most often used in combination with a partition $\mathscr{B}$ of Ω. In that case, the conditional information along the elements of the partition can be aggregated in a single set as we show later on in (16).

8. Conglomerability: the Non-Causal STP

We consider a temporal setting resembling a proposal of ours done with different aims in some past work [22, Section 3]. In such a setting, we establish our (κ-)coherent set $\mathscr{D}$ at present time. At some future point, an event B occurs from a partition $\mathscr{B}$ of possible events. When B occurs we are given the chance to reconsider our assessment so as to define our new (κ-)coherent set of assessments $\mathscr{D}^B$, which takes into account the fact that B has occurred.

Now, suppose that we know the partition $\mathscr{B}$ of possible events already at present time. Thus we know that one of those events will later occur. If we assume that our setting is one of 'perfect information' (no information other than B will be revealed from present time to the occurrence of the event), we might well be prepared, already at present time, to define sets $\mathscr{D}^B$, for all $B \in \mathscr{B}$, equal to the conditional sets $\mathscr{D}|B$, and commit to adopting the one that corresponds to the event B occurring, right after it occurs. In [24, Section 6], we have argued that this approach is justified if we establish our assessments $\mathscr{D}$ in a 'reliable' way, which means never making statements stronger than the evidence allows; and in addition, that our perception of gambles' values stays the same during this process, which is something that we called 'perfect isolation'.

In summary, under reliability of the assessments' procedure, as well as perfect information and isolation, it is sensible to assume that we commit, for the future, to our conditional sets already at present time.

Let us then represent our overall future commitment by the so-called 'conglomerable natural extension' of the conditional sets:

$$\mathscr{D}|\mathscr{B} := \left\{ \sum_{B\in\mathscr{B}} Bf : (\forall B)\ Bf \in \mathscr{D}|B \cup \{0\} \right\} \setminus \{0\}. \quad (16)$$

Taking into account that only one of the events B will occur, the set $\mathscr{D}|\mathscr{B}$ will correctly reduce to $\mathscr{D}|B$ in the future, thus representing the commitments that get enforced after B. But since whatever $\omega \in \Omega$ will occur in the future, there will be a $B \in \mathscr{B}$ that contains it, and since we commit to Bf, the actual result is that we will accept $f(\omega)$, whatever ω will occur; whence already now we are committed to accept f.

This makes us wonder about the kind of relation that exists between $\mathscr{D}$ and $\mathscr{D}|\mathscr{B}$. For, after all, both these sets concern commitments that we do at present time. And for this reason, they should not be inconsistent. As we have argued in [22, Section 6.4],[5] in this case we should require that present commitments $\mathscr{D}$ (strongly) cohere with future ones $\mathscr{D}|\mathscr{B}$, which in practice means that the following inclusion must hold:

$$\mathscr{D}|\mathscr{B} \subseteq \mathscr{D}. \quad (17)$$

This is just non-causal STP. Note that we have not used additivity (D3) to derive it: let us remark, indeed, that we are working in a very general setting of nonlinear utility scales, thanks to the generality of κ; the sums over the partition $\mathscr{B}$ that define $\mathscr{D}|\mathscr{B}$ are only a consequence of the fact that the events B lead to mutually exclusive coherent sets in the future.

Finally, note that the argument leading to (17) is not bound to the cardinality of $\mathscr{B}$, which can be (also uncountably) infinite; this means that the present discussion justifies the use of $\mathscr{B}$-conglomerability also in the generalised κ-desirability setting. In fact, we shall refer to (17) just as *conglomerability* from now on.

Example 1 *Take $\Omega := \{1,2,3,4\}$ and the closure operator κ given by:*

$$k(\mathscr{D}) := \begin{cases} \mathscr{D} \ \text{if}\ (\forall f \in \mathscr{D}) |\{\omega \in \Omega : f(\omega) < 0\}| \leq 1 \\ \mathscr{L} \ \text{otherwise}. \end{cases}$$

5. It it not difficult to prove that the results in that section hold also for κ-coherent sets of gambles.

Let $\mathscr{B} := \{B, \bar{B}\}$, with $B := \{1,2\}$, and

$$\begin{aligned} \mathscr{D}|B &:= \{f \in \mathscr{L} : f(2) > 0, f(3) = f(4) = 0\}, \\ \mathscr{D}|\bar{B} &:= \{f \in \mathscr{L} : f(4) > 0, f(1) = f(2) = 0\}, \\ \mathscr{D} &:= \mathscr{D}|B \cup \mathscr{D}|\bar{B} \cup \mathscr{L}^+. \end{aligned}$$

Then $\kappa(\mathscr{D} \cup \mathscr{L}^+) = \mathscr{D}$, so this set is ($\kappa$)-coherent. However, by (16), $\mathscr{D}|\mathscr{B}$ contains the gamble $(-1,1,-1,1)$, so $\kappa(\mathscr{D}|\mathscr{B}) = \mathscr{L}$ and $\mathscr{D}|\mathscr{B}$ incurs partial loss. As a consequence, $\mathscr{D}$ and $\mathscr{D}|\mathscr{B}$ are strongly temporally incoherent.

In order to solve the incoherence, we should impose (17). But if we do that, then $\mathscr{D}$ would become incoherent. So there is no possibility to solve the incoherence.

This tells us that our assessments $\mathscr{D}$ are irrational under operator κ, if we also ought to maintain strong temporal coherence; as a consequence $\mathscr{D}$ must be rejected. $\Diamond$

This example illustrates a few important traits of the problem under examination. First, with a nonlinear operator κ, the conglomerable natural extension can incur partial loss, unlike with posi. For the same reason, the conglomerable natural extension may not automatically belong to our assessments $\mathscr{D}$ in the case of finite partitions $\mathscr{B}$. Stated differently, with a nonlinear operator κ, conditioning, under the common temporal interpretation, is not automatically well behaved: we see the rise of problems that remind us of the issues that conglomerability originates in probability, while actually becoming more severe.

The positive finding, in our view, is that the temporal framework we have sketched maintains its validity no matter the nonlinearity of the closure operator. It leads us to conglomerability in (17) while being more primitive.

9. Causal STP

Our goal in this section is to prove that the causal STP may hold also in a setting made of nonlinear utility scales. We know that in the linear case, as in Pearl's [13, Theorem 6.1.1], the proof can be traced back to total probability plus the causal independence condition (3).

We can try to follow the same path relying on the generalisation of total probability to desirability:

Definition 13 (Marginal extension) *Let $\mathscr{B}$ be a partition of Ω. Let $\mathscr{D}_{\mathscr{B}}$ be a (marginal) set coherent relative to $\mathscr{L}_{\mathscr{B}}$; and $\mathscr{D}|\mathscr{B}$ be the set originated by (16) via coherent conditional sets. The* marginal extension *of $\mathscr{D}_{\mathscr{B}}$ and $\mathscr{D}|\mathscr{B}$ is defined as*

$$\mathscr{E}_\kappa(\mathscr{D}_{\mathscr{B}} \cup \mathscr{D}|\mathscr{B}).$$

Note that the marginal extension contains $\mathscr{D}|\mathscr{B}$, thus satisfying the rationality requirement of conglomerability. And yet, since $\mathscr{D}|\mathscr{B}$ may incur partial loss (see Example 1), the marginal extension may fail to be coherent. But coherence can well hold: it does, for example, if the closure operator satisfies

$$(\forall \mathscr{D})\ \kappa(\mathscr{D}) \subseteq \mathrm{posi}(\mathscr{D}).$$

So our first assumption is that the marginal extension yields coherent sets. This is the counterpart of Pearl's use of total probability—which is coherent thanks to $\kappa = \mathrm{posi}$.

To introduce the second assumption we have to start describing the causal-STP framework: our initial assessments (prior to actions) are represented by a coherent set $\mathscr{D}$. We have to decide between taking action a and action b. The (causal) independence assumption is that actions will not affect our marginal set $\mathscr{D}_{\mathscr{B}} := \mathscr{D} \cap \mathscr{L}_{\mathscr{B}}$. They are instead free to arbitrarily change our conditional assessments into $\mathscr{D}_a|\mathscr{B}$ and $\mathscr{D}_b|\mathscr{B}$, respectively.

As a consequence, if we take action a, our assessments will become $\mathscr{D}_a := \mathscr{E}_\kappa(\mathscr{D}_{\mathscr{B}} \cup \mathscr{D}_a|\mathscr{B})$; similarly, if we take b, we will get $\mathscr{D}_b := \mathscr{E}_\kappa(\mathscr{D}_{\mathscr{B}} \cup \mathscr{D}_b|\mathscr{B})$.

Now assume that for a certain gamble f we have that

$$Bf \in \mathscr{D}_a|\mathscr{B} \text{ and } Bf \notin \mathscr{D}_b|\mathscr{B}$$

for all $B \in \mathscr{B}$. Stated differently, if B is observed, then action a is better than b w.r.t. f, and this irrespective of the $B \in \mathscr{B}$ we may consider. The causal STP means then that there is no real number μ such that

$$f - \mu \notin \mathscr{D}_a \text{ while } f - \mu \in \mathscr{D}_b,$$

i.e., that preferences cannot be reversed in the global model.

Our next result gives a sufficient condition for STP:

Theorem 14 *Let κ be a closure operator such that any coherent set of desirable gambles is closed under* dominance*:*

$$f \geq g \in \mathscr{D} \Rightarrow f \in \mathscr{D};$$

and assume that its associated marginal extension operator is such that for every $f \in \mathscr{E}_\kappa(\mathscr{D}_{\mathscr{B}} \cup \mathscr{D}|\mathscr{B})$ there are $g_1 \in \mathscr{D}_{\mathscr{B}} \cup \{0\}, g_2 \in \mathscr{D}|\mathscr{B} \cup \{0\}$ such that

$$f \geq g_1 + g_2. \tag{18}$$

Then κ satisfies the causal STP.

Proof Consider a gamble f such that $Bf \in (\mathscr{D}_a|\mathscr{B}) \setminus (\mathscr{D}_b|\mathscr{B})$ for every $B \in \mathscr{B}$. It follows by (17) that $f \in \mathscr{D}_a|\mathscr{B}$ and as a consequence also to $\mathscr{D}_a$. Now, if for $\mu \geq 0$ it holds that $f - \mu \in \mathscr{D}_b$, then by (18) there are $g_1 \in \mathscr{D}_{\mathscr{B}} \cup \{0\}, g_2 \in \mathscr{D}_b|\mathscr{B} \cup \{0\}$ such that $f - \mu \geq g_1 + g_2$. As a consequence, $B(f-\mu) \geq Bg_1 + Bg_2$ for every $B \in \mathscr{B}$. If $g_1(B) \geq 0$, we deduce that $B(f-\mu) \geq Bg_2 \in \mathscr{D}_b|\mathscr{B}$, and, since we are assuming that a coherent set of gambles is closed under dominance, it follows that $B(f-\mu) \in \mathscr{D}_b|\mathscr{B}$ and therefore also $Bf \in \mathscr{D}_b|\mathscr{B}$, a contradiction. From here it follows that it should be $g_1(B) < 0$ for every $B \in \mathscr{B}$, but this contradicts the coherence of the set $\mathscr{D}_{\mathscr{B}}$. We conclude that $f - \mu \notin \mathscr{D}_b$ for any $\mu \geq 0$ and as a consequence STP holds. ■

The point of this theorem is to show that the causal STP does not depend on linearity (posi) to be true. The coherence of marginal extension and Pearl's causal independence condition suffice to prove it in quite a broad contest: e.g., it is not difficult to show that the assumptions of the theorem are satisfied by the following closure operators:

- $\kappa_1(\mathscr{D}) = \mathrm{posi}(\mathscr{D})$.
- $\kappa_2(\mathscr{D}) = \{g \geq f \text{ for some } f \in \mathscr{D}\}$.
- $\kappa_3(\mathscr{D}) = \{g \geq \lambda f \text{ for some } f \in \mathscr{D}, \lambda > 0\}$.
- $\kappa_4(\mathscr{D}) = \{\sum_{i=1}^n f_i \text{ for some } f_1, \dots, f_n \in \mathscr{D}, n \in \mathbb{N}\}$.

Out of these, κ_1 is the closure operator used in traditional (linear) coherence; κ_2 is the closure operator associated with 2-convexity [16]; while κ_3 is related, but not equivalent, to 2-coherence, as showed also in [16].[6]

On the other hand, it is possible to show that the sufficient condition (18) for causal STP in Theorem 14 is not necessary, and also that there are closure operators for which causal STP is *not* satisfied, i.e., that the condition is not trivial.

10. Conclusions

We have shown that in a non-causal setting, under non-linear utility scales, STP coincides with conglomerability; and that it can be justified via considerations of temporal coherence. This part appears to be particularly relevant to give conglomerability the status of an axiom to adopt when logic is concerned with the 'updating' of one's assessments.

Then we have shown that the causal STP can be derived mimicking Pearl's proof while working in quite a broader setting made of nonlinear utility scales and imprecision. This tells us that the causal STP is more primitive than probability, as it can follow just from general consistency properties in logic (κ-coherence) and a causal independence condition.

Our proof gives a sufficient condition. It would be interesting to explore the limits of validity of the STP, thus yielding a necessary condition too.

Acknowledgments

We acknowledge the financial support of project PGC2018-098623-B-I00.

References

[1] F. J. Anscombe and R. J. Aumann. A definition of subjective probability. *The Annals of Mathematical Statistics*, 34:199–2005, 1963.

6. The difference is that the axiomatisation of 2-coherence in terms of desirability requires in addition that the sum of two desirable gambles must have a positive supremum.

[2] R. Aumann, S. Hart, and M. Perry. Conditioning and the sure-thing principle. Discussion paper series, The Federmann Center for the Study of Rationality, the Hebrew University, Jerusalem, 2005. URL http://www.ma.huji.ac.il/hart/papers/stp.pdf.

[3] P. Berti, E. Miranda, and P. Rigo. Basic ideas underlying conglomerability and disintegrability. *International Journal of Approximate Reasoning*, 88(C): 387–400, 2017.

[4] C. R. Blyth. On Simpson's paradox and the surething principle. *Journal of the American Statistical Association*, 67(338), 1972.

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

[6] G. de Cooman and M. Zaffalon. Updating beliefs with incomplete observations. *Artificial Intelligence*, 159(1–2):75–125, 2004.

[7] B. de Finetti. Sulla proprietà conglomerativa delle probabilità subordinate. *Rendiconti del Reale Instituto Lombardo*, 63:414–418, 1930.

[8] B. de Finetti. Sull'impostazione assiomatica del calcolo delle probabilità. *Annali Triestini dell'Università di Trieste*, 19(2a):29–81, 1949. Reprinted in [9, Section 5.31].

[9] B. de Finetti. *Probability, Induction and Statistics*. Wiley, London, 1972.

[10] A. Gibbard and L. Harper. Counterfactuals and two kinds of expected utility. In W. L. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs, Conditionals, Belief, Decision, Chance, and Time*, pages 153–169. D. Reidel, 1981.

[11] P. Grünwald and J. Y. Halpern. Updating probabilities. *Journal of Artificial Intelligence Research*, 19:243–278, 2003.

[12] R. Jeffrey. The sure thing principle. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume 2, pages 719–730. The University of Chicago Press, 1982.

[13] J. Pearl. *Causality*. Cambridge University Press, second edition, 2009.

[14] J. Pearl. Comment: understanding Simpson's paradox. *The American Statistician*, 68(1):8–13, 2014.

[15] J. Pearl. The sure-thing principle. *Journal of Causal Inference*, 4(1):81–86, 2016.

[16] R. Pelessoni and P. Vicig. 2-coherent and 2-convex conditional lower previsions. *International Journal of Approximate Reasoning*, 77:66–86, 2016.

[17] E. Quaeghebeur. Desirability. In T. Augustin, F. Coolen, G. de Cooman, and M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 1–27. Wiley, 2014.

[18] L. J. Savage. *The Foundations of Statistics*. Dover, New York, 1972. Second revised edition, first published 1954.

[19] E. Simpson. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B*, 13:238–241, 1951.

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

[21] M. Zaffalon and E. Miranda. Conservative inference rule for uncertain reasoning under incompleteness. *Journal of Artificial Intelligence Research*, 34:757–821, 2009.

[22] M. Zaffalon and E. Miranda. Probability and time. *Artificial Intelligence*, 198(1):1–51, 2013.

[23] M. Zaffalon and E. Miranda. Axiomatising incomplete preferences through sets of desirable gambles. *Journal of Artificial Intelligence Research*, 60:1057–1126, 2017.

[24] M. Zaffalon and E. Miranda. Desirability foundations of robust rational decision making. *Synthese*, 2018.