A graph model for probabilities of nested conditionals

Anna Wójtowicz¹ · Krzysztof Wójtowicz¹

Accepted: 12 February 2021 / Published online: 30 July 2021
© The Author(s) 2021

Abstract We define a model for computing probabilities of right-nested conditionals in terms of graphs representing Markov chains. This is an extension of the model for simple conditionals from Wójtowicz and Wójtowicz (Erkenntnis, 1–35. <https://doi.org/10.1007/s10670-019-00144-z>, 2019). The model makes it possible to give a formal yet simple description of different interpretations of right-nested conditionals and to compute their probabilities in a mathematically rigorous way. In this study we focus on the problem of the probabilities of conditionals; we do not discuss questions concerning logical and metalogical issues such as setting up an axiomatic framework, inference rules, defining semantics, proving completeness, soundness etc. Our theory is motivated by the possible-worlds approach (the direct formal inspiration is the Stalnaker Bernoulli models); however, our model is generally more flexible. In the paper we focus on right-nested conditionals, discussing them in detail. The graph model makes it possible to account in a unified way for both shallow and deep interpretations of right-nested conditionals (the former being typical of Stalnaker Bernoulli spaces, the latter of McGee’s and Kaufmann’s causal Stalnaker Bernoulli models). In particular, we discuss the status of the Import-Export Principle and PCCP. We briefly discuss some methodological constraints on admissible models and analyze our model with respect to them. The study also illustrates the general problem of finding formal explications of philosophically important notions and applying mathematical methods in analyzing philosophical issues.

✉ Anna Wójtowicz
amwojtow@uw.edu.pl
Krzysztof Wójtowicz
wojtow@uw.edu.pl

¹ Department of Philosophical Logic, Faculty of Philosophy, University of Warsaw, Krakowskie Przedmieście 3, 00-927 Warszawa, Poland

Keywords Probability of conditionals · Conditional probability · PCCP · Nested conditionals · Subjunctive conditionals · Indicative conditionals · Markov graph · Stalnaker Bernoulli model

1 Introduction

Conditional sentences are a hard nut to crack. Analysis fails in terms of propositional calculus with the conditional as the material implication. It is obvious that conditionals are not ordinary factual sentences—their truth value and truth conditions are a notorious problem. It is also not clear what the proper logic of conditionals should be: what are the appropriate axioms and what are the rules of inference?

Our approach in the paper is to focus on the probabilities of conditionals. We propose a kind of game (process) semantics, so conditions under which the respective game is won (lost) have to be specified. This means that we assume that it is reasonable to speak of circumstances in which the status of the conditional is settled, and in particular that the notion of truth conditions is essential in our account. The construction of the appropriate graphs (and the corresponding probability spaces) relies on this assumption.¹

Adams took the first steps towards analysis of the probability of conditionals and made an important pioneering contribution, **defining** the probability of a conditional $A \rightarrow B$ as $P(B|A)$ (Adams 1965, 1970, 1975, 1998). This stipulation seems very natural for many examples. Indeed, consider the roll of a fair die and the conditional: *If it is even, then it is a six*. Intuitively, the probability of this conditional is $1/3$, which is exactly the conditional probability $P(It is a six | It is even)$.²

Adams' thesis (also called 'Stalnaker's thesis') that probabilities of simple conditionals can be defined as conditional probabilities is attractive as it resolves many problems and is coherent with intuitive judgement for many examples. However, the applicability of this approach is limited and the triviality results of Lewis (1976) (see Hajek 2011 for elegant generalizations) indicate that this is a fundamental problem.

Simple conditionals are only the first step in the enterprise: we want to deal with nested conditionals, conjoined conditionals, and perhaps also with other combinations. However, Adams declared that his approach does not extend to compound conditionals: "we should regard the inapplicability of probability to compounds of conditionals as a fundamental limitation of probability, on a par with the inapplicability of truth to simple conditionals" (Adams 1975, 35).³ Our approach is different: we think that it is possible to **compute** the probabilities of conditionals (simple and compound) by modeling them within appropriate probability spaces

¹ We are indebted to the anonymous referees for pointing out that this issue requires clarification, in particular for encouraging us to make more explicit claims concerning the status of truth conditions in our model.

² "What is the probability that I throw a six if I throw an even number, if not the probability that if I throw an even number, it will be a six?" (van Fraassen 1976, 273).

³ Also Edgington (1991, 1995) and Gibbard (1981) claim that conditionals are not factual sentences.

(the technical devices used are Markov graphs, which represent Markov chains). Therefore, our approach differs from Adams' in two important respects: (i) we believe that it is possible to apply the notion of probability to compound conditionals; (ii) we believe that the notion of truth conditions is applicable to conditionals. They are usually much more complex than for conditional-free sentences, as we need game scenarios in our approach, which are formally sequences of possible worlds, not worlds *simpliciter*.⁴

The intuitive judgements of the probability of conditionals are unproblematic for a limited number of cases only, which shows that we have to be very cautious when speaking about probability. Matters become rather complicated when we take the context into account. Using the notion of probability in a purely intuitive sense might lead to misunderstandings and contradictions—contrary to Laplace's claim, we are not good intuitive statisticians, as many empirical results show. So it is necessary to construct a formal model which will give a precise meaning to our understanding of the notion of the probability of conditionals. This will enable us to move from purely qualitative to quantitative analysis. This problem has already been discussed extensively and there are many results available—we present a short overview in Sect. 3.

The aim of this article is to present a formal model which allows the probabilities of conditionals to be computed; here we focus on right-nested (compound) conditionals, but the model can be applied to a wider class of examples. This model is an extension of the model for simple conditionals defined in Wójtowicz and Wójtowicz (2019). In a nutshell, the idea is to compute the probability of conditionals by means of Markov chains, which have a very simple, intuitive and natural illustration in the form of graphs. These graphs exhibit the “inner dynamics” of the conditional and allow to identify its (usually context-sensitive) interpretation. We associate a graph G_α (representing a Markov chain) with the conditional α in question, taking the respective interpretation into account when necessary. This graph is connected in a canonical way with a probability space S_α , in which the conditional α is represented as an event, therefore its probability can be computed in a formally sound way. In this way our intuitive assessments are given a precise formal underpinning, and our intuitions can be elucidated by formal means.

In our opinion, the graph model fulfills its role better (at least in some respects) than other models known from the literature because it is formally very simple, intuitive and more flexible.

The structure of the article is as follows:

In Sect. 2, *Formal model for conditionals* we give a short analysis of the requirements an adequate model has to meet—in particular, which problems we expect to be solved by the graph model.

⁴ We do not discuss the problem of low/high assertability. Indeed, in Adams (1965) the term “assertability” is used. As Khoo observed, the idea that assertability is also governed by conditional probability is widespread (Khoo 2015, 2); see also (Hajek 2012) for discussion.

In Sect. 3, *Right-nested conditionals*, we discuss the problem of the interpretation of nested conditionals and indicate that there are (at least) two context-dependent intuitive interpretations.

In Sect. 4, *Some classic results*, we recall some of the results, focusing on the results obtained with the use of Stalnaker Bernoulli space by van Fraassen and Kaufmann; some of McGee's results are also presented.

In Sect. 5, *The graph model for a simplified case*, we present the idea of modeling conditionals in terms of graphs, taking a simple conditional $A \rightarrow B$ as an example. The section also contains a presentation of the needed notions of Markov graph (chains) theory, as this is the formalism we use in our model.

In Sect. 6, *The graph model for right-nested conditionals*, we present the graph-theoretic representation for both interpretations of right-nested conditionals (i.e. the shallow and the deep interpretation); we compute the corresponding probabilities and show that they agree with the results known from the various models of Kaufmann and McGee.

In Sect. 7, *Multiple-(right)-nested conditionals*, we show how to generalize our approach to the case of more complex nested conditionals of the form $A_1 \rightarrow (A_2 \rightarrow \dots \rightarrow (A_n \rightarrow B) \dots)$. The graphs for both the shallow and the deep interpretations are obtained in a straightforward way and make the computations very easy. We also discuss the problem of a longer conditional where the conditional operator is interpreted in a non-uniform, "mixed" way, i.e. as shallow for some occurrences and as deep for others. We discuss the problem in more detail for conditionals of length four, i.e. of the form $A_1 \rightarrow (A_2 \rightarrow (A_3 \rightarrow B))$ as they are still natural, but complex enough for interesting phenomena to occur.

In Sect. 8, *Formal models for conditionals revisited* we show that our model solves some of the specific issues (the status of IE and PCCP), that it fulfills the methodological criteria formulated in Sect. 1 and that according to these criteria it fares better than other available models. A direct comparison with the Stalnaker Bernoulli model is also made (7.3).

Section 9 is a short *Summary*.

2 Formal models for conditionals

Formal models of conditionals which allow the use of the concept of probability in a mathematically correct way and without the risk of falling into paradoxes are built to clarify and sharpen the intuitions of ordinary language users. There are numerous "test sentences" for which the probability judgements are intuitively obvious (as in the *If it is even, it is a six* example), and the model should respect them. After all, we are talking about **conditionals** and not about some exotic operator which we are trying to introduce into the language. Generally speaking, we want to "translate" the results of qualitative analysis into quantitative language, hoping that the formal models will allow vague notions to be explicated, shed light on the state of the art, and inspire discussion by presenting formal results (which might happen to be

technical counterparts of intuitive claims).⁵ In particular we expect them to provide a better understanding of (at least some of) the debated issues. We also think, that it is useful to keep in mind possible methodological virtues of the model and we consider generalizability and simplicity to be important in this context (2.2.1 and 2.2.2). A resume and discussion of these issues is to be found in Sect. 7. In the paper, letters *A*, *B*, *C*... symbolize non-conditional propositions. We will usually use italics for propositions, and letters without italics for the corresponding events.

2.1 Two important principles

Two important principles discussed in the literature devoted to probability of conditionals are: (i) the Import-Export Principle; (ii) PCCP. We remind them briefly (2.1.1 and 2.1.2), we hope this will be useful for the readers, as these principles are discussed in the technical fragments of the text.

2.1.1 The Import-Export Principle

An important problem (indirectly related to the issue of generalization) is the status of the Import-Export (IE) Principle. This is very attractive as it simplifies matters enormously: right-nested conditionals $A \rightarrow (B \rightarrow C)$ are reduced to simple conditionals $(A \wedge B) \rightarrow C$. Its consequence is accepting that $P(A \rightarrow (B \rightarrow C)) = P((A \wedge B) \rightarrow C)$.⁶

On one hand, such a rule is extremely convenient because computing the probabilities is simpler, and generalization to a wide class of conditionals like $A \rightarrow (B \rightarrow (C \rightarrow D))$ is straightforward. On the other hand, there are natural examples where $P(A \rightarrow (B \rightarrow C)) \neq P(B \rightarrow (A \rightarrow C))$, but this means that the Import-Export Principle is violated. A good model should identify these cases and explain under which presuppositions it is justified to accept the IE Principle. In particular it is important to discuss it in the context of the possible interpretations (deep versus shallow) of right-nested conditionals. We show that in our model IE can be proved for the deep interpretation (also for the multiple-right-nested conditionals), and that it is generally not true under the shallow interpretation.

2.1.2 PCCP

Is $P(A \rightarrow B) = P(B|A)$? Namely, is the probability of the conditional equal to the conditional probability? (We will use the familiar acronym PCCP henceforth.) It would be nice to have PCCP as it gives a very simple way of computing

⁵ The general issue of context modeling arises, in particular the question whether “fine-tuning of context” is available within the model(s). There are also many interesting philosophical problems involved, for instance: (i) the possible influence of the interpretation of the notion of probability on the model; (ii) the metaphysical status of possible worlds (and their sequences); (iii) the explanatory value of the model(s). However, the detailed analysis of these issues exceeds the scope of the present study (preliminary results are discussed in Wójtcowicz and Wójtcowicz 2019, 2021).

⁶ The thesis, that logically equivalent propositions have identical probabilities is widely accepted, so if we accept the equivalence $(A \rightarrow (B \rightarrow C)) \Leftrightarrow ((A \wedge B) \rightarrow C)$ (i.e. IE), the equation $P(A \rightarrow (B \rightarrow C)) = P((A \wedge B) \rightarrow C)$ follows.

probabilities. Also, we have strong intuitions linking the probability of conditionals with the conditional probability, as the classic quotation from Ramsey states.⁷ But on the other hand, we have Lewis' and Hajek's triviality results (Lewis 1976; Hajek 2011, 2012).⁸ So, the formal model for conditionals must identify the relationships between the probability of the conditional and the respective conditional probability in order to shed light on the status of (the appropriate version of) PCCP.⁹ We prove $P(A \rightarrow B) = P(B|A)$ for the simple conditional $A \rightarrow B$, and identify and prove the generalizations of PCCP for the deep and the shallow interpretations of the right-nested conditional $A \rightarrow (B \rightarrow C)$.

2.2 Methodological aspects

We consider generalizability and simplicity to be important features of formal models for conditionals. Generalizability (and flexibility) might be viewed to be good indicators that the model will also be applicable in cases which were not a direct motivation for its creation. The simplicity of the model makes it applicable.

2.2.1 Generalizability and flexibility

Regardless of whether we analyze simple conditionals $A \rightarrow B$, nested conditionals $A \rightarrow (B \rightarrow C)$ and $(A \rightarrow B) \rightarrow C$, conjoined conditionals $(A \rightarrow B) \wedge (C \rightarrow D)$, or conditional conditionals $(A \rightarrow B) \rightarrow (C \rightarrow D)$, the conditional operator $\rightarrow$ which occurs in them is, **in principle**, similar. So, if we have a definite method of building a model for probabilities of simple conditionals $A \rightarrow B$, a natural hypothesis is that this method will also be applicable in cases in which complex formulas will stand for A and B . If there are any contexts in which the operator $\rightarrow$ is given different interpretations, our model should be able to represent this fact.

2.2.2 Formal simplicity

An attractive model needs to be “user friendly” as modeling conditionals should be as simple as possible. If we think of the problem also in practical terms—for instance, if we are interested in analyzing decision-making processes (based on evaluating the probabilities of conditionals)—it is important to have a method which is of practical as well as purely theoretical interest. In short, formal treatment of more complex cases must be feasible.

⁷ “If two people are arguing ‘If p will q?’ and both are in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q... We can say that they are fixing their degrees of belief in q given p.” (Ramsey 1929, 247).

⁸ For empirical studies concerning evaluating the probabilities of conditionals see for instance (Douden and Verbrugge 2013). Their findings seem to indicate that $P(A \rightarrow B) = P(B | A)$ does not hold.

⁹ In Wójtowicz and Wójtowicz (2021) we give an independent argument in favor of PCCP based on a diachronic Dutch Book reasoning.

3 Right-nested conditionals

In ordinary life we often utter sentences which have the structure of a nested conditional $A \rightarrow (B \rightarrow C)$. Examples are abundant:

- a. *If the match is wet, then it will light if you strike it.*
- b. *If John had broken his leg while skiing in Davos in 2017, it would hurt if he went skiing.*
- c. *If John takes medication A, then if he takes medication B, he will have an allergic reaction.*¹⁰

The first example is known from the literature. The last two sentences are our toy examples in this section. All these conditionals are right-nested, i.e. they have the same form $A \rightarrow (B \rightarrow C)$, but the evaluation of their probabilities might depend on the interpretation.

In (a) we quite obviously mean the same match, which is wet, which is struck, and which lights or not: a unity of action, place and time (like in an ancient tragedy). Borrowing terminology from Kaufmann (2015), we call this interpretation **deep**. However, in (b) we obviously do not mean the claim that the leg would have hurt if John had continued skiing **with his leg broken in Davos in 2017**. The sentence does not apply to this particular unfortunate skiing holiday. This is obvious in light of the common knowledge concerning broken legs and there is no need even to mention it. What we mean is rather that this accident would have some consequences for John's later life: speaking informally, John's life would be "redirected" to a route where the probability of pain while skiing is greater. We mean subsequent skiing holidays, not 2017 in Davos. So, the context imposes a different interpretation of this sentence in a natural way: following Kaufmann (2015) we call it **shallow**.

For the last sentence (c), the interpretation is not so obvious. (We will use the natural abbreviation MED for (c).) To fix attention, we think in terms of days, i.e. we perform the appropriate observations (*Which medications are taken? What is the reaction?*) within the span of a single day. When interpreting MED, we analyze either whether

- i. taking B **at the same moment** (i.e. the same day) as taking A will cause the allergic reaction, or whether
- ii. taking B **any time** after taking A will cause the allergic reaction.

These different interpretations should have their counterparts in two different sets of truth conditions. Imagine two agents, Alice and Bob, who observe consecutive days in John's life (i.e. whether he took A, whether he took B, and whether an allergic reaction occurs). Alice thinks that MED will turn out to be true, but Bob claims the opposite, so they make a bet. We can say that they are going to play the MED game.

Assume that they have been watching John since Monday.

¹⁰ In a different formulation: *If John takes medication A, then he will have an allergic reaction if he takes B.*

Monday: John does **not** take A but takes B and has an allergic reaction. Alice and Bob agree that the bet has not been decided as the antecedent A has not occurred. Indeed, John's allergic reaction is irrelevant with respect to the relationship expressed in MED. The game continues.

Tuesday: John takes A but does not take B. Alice and Bob agree that the bet has not been decided: the internal antecedent has not occurred (and it does not matter whether he has had an allergic reaction). So, the game continues.

Wednesday: John does not take A but takes B and has an allergic reaction. This is exactly the situation in which Alice's and Bob's opinions differ:

Bob: *Unfortunately, you win! Indeed, John took A (remember that was yesterday), and here is the winning combination, i.e. "medication B + an allergic reaction"!*

Alice counters: *I'm afraid not! The bet was about John having an allergic reaction **provided he took both the medications A and B on the same day**—not just some allergic reaction of an unknown etiology that incidentally occurred after John took **only** medication B on that day. This is obvious from the context and means that the bet is undecided: we have to wait until John **takes A and B on the same day**, and we will see what happens then.*

Who is right? This is a matter of interpretation based on some (tacit) presuppositions.¹¹ Alice accepts the deep interpretation (i), whereas Bob accepts the shallow interpretation (ii). We think that—at least in the case of some medications—both the interpretations might be legitimate depending on the context, in particular on the medical (biochemical) findings and also on the features of the individual (John).

Anyway, if we are going to make a bet on MED, we must first define the terms (in particular, when the game is won and lost), i.e. we have to agree on an interpretation of the conditional. The crucial difference is what happens after John has taken medication A but not medication B, therefore the bet has not yet been settled. Alice and Bob give different answers:

Alice (deep interpretation): the game is restarted. We have to observe John's behavior anew, and wait until a day when John takes both medications A and B (and the history of his taking A on previous days has no influence on the probabilities of taking the medications and the reaction). If the allergic reaction occurs, I win; if it does not, I lose.

Bob (shallow interpretation): the game proceeds on different terms as we now treat the condition of taking medication A as fulfilled "forever". If John's taking B is

¹¹ For instance, A might be known **not** to cause any long-term allergic reactions. In this case the bet (in order not to be trivial) must be about the combination A + B, i.e. about the deep case. Or it might be the other way around: A + B taken together always cause an allergic reaction. In this case the non-trivial hypothesis must concern the long-term consequences of taking A (i.e. the shallow interpretation).

followed by an allergic reaction, you will win (regardless of whether John has taken A that day). If it is not, you will lose.¹²

We will turn to the problem of computing the probabilities of MED later, when the necessary technical tools have been introduced. However, it is intuitively clear that depending on the interpretation, the probabilities might differ vastly.¹³ It might well be that two rational agents assign different probabilities to the same conditional, even if they agree on the probabilities of the factual (“atomic”) sentences. This can be a result of interpreting them in a different way (e.g. deep versus shallow) or relying on some assumptions concerning the context. We claim that the presented model explains this phenomenon by (formally) identifying these interpretations.¹⁴

4 Some classic results

Adams defines the probability of a conditional $A \rightarrow B$ by setting $P(A \rightarrow B) := P(B|A)$. This simplifies matters, but this stipulation rests on quite controversial assumptions. For instance it is (by definition) necessary to accept PCCP as a general principle (but this thesis is disputed, and arguments against it cannot be neglected). This approach cannot be easily extended to compound conditionals. But this means, that P might behave not like a genuine probability function, as it might happen, that $P(\alpha)$ and $P(\beta)$ are defined, but $P(\alpha \rightarrow \beta)$ is not.¹⁵

The possible worlds semantics has become standard in treating modalities, so it is not surprising that there have been attempts to account for the problem of conditionals (both their truth conditions and probabilities) in these terms (the classic papers are, for instance, Stalnaker 1968, 1970; Stalnaker and Thomason 1970). The general idea is familiar: if $[\alpha]$ is the set of worlds corresponding to the proposition α

¹² So, Alice’s interpretation is the counterpart of the wet match example; Bob’s interpretation is the counterpart of the Davos example: you remain forever a person who has (ever) had a broken leg. Consider also a modified version of the Davos example: *If John had burned his leg while skiing in Davos in 2017, it would hurt if he went skiing*. It is perfectly reasonable to assume that you can continue skiing with a burn on your leg provided it is not too bad and the pain is not very annoying. But if this is a serious injury and you have to stay in hospital, then of course skiing is not possible. So, the parameter upon which the interpretation depends is the information concerning the kind of injury.

¹³ For instance, if there is **no** day when John takes both A and B, then it will never be possible to settle the bet under the deep interpretation because the circumstances in which the conditional can be decided will never occur. However, under the shallow interpretation the bet might be won very quickly: the first day being $A \wedge \neg B \wedge \neg R$ (R is for Reaction), and the second day being $\neg A \wedge B \wedge R$.

¹⁴ Bradley (2012) discusses the problem of explaining a system of beliefs concerning the probabilities of conditionals (as well as some logical relations between them) in terms of producing a possible worlds model which fits that system. We think that our approach fulfills this role.

¹⁵ If we accept the Import-Export Principle, then a right-nested conditional $A \rightarrow (B \rightarrow C)$ can be replaced by the (simple) conditional $(A \wedge B) \rightarrow C$. In this case its probability of $A \rightarrow (B \rightarrow C)$ equals $P(C|AB)$ —which is well defined, as AB (in the paper we denote the intersection of events A and B by AB) is an event in the sample space. However, it is not clear how other conditionals can be handled, for instance left-nested conditionals $(A \rightarrow (B \rightarrow C))$. It is generally true that if α is an event and $P(\alpha) > 0$ then $P(\cdot|\alpha)$ is a probability function. But conditionals need not be events in the original sample space, so the expression $P(\cdot|\alpha)$ does not always make sense; for instance $P(\cdot|(A \rightarrow B))$ is not defined. (We thank the anonymous referee for helping us to clarify these matters).

(i.e. in which α is true), then the probability of the proposition α is given by $P([\alpha])$ (where P is the probability distribution defined on the set of possible worlds). This is natural for propositions not containing conditionals, but it is far from obvious how to apply it to conditionals. In general, it is not clear which set of possible worlds should be assigned to $A \rightarrow B$ (i.e.—what is $[A \rightarrow B]$). There have been numerous attempts to overcome this problem. One proposal is to interpret conditionals as random variables which are given appropriate values at different possible worlds, and the probability of the conditional $A \rightarrow B$ (or generally, its semantic value) is defined as the conditional expected value (Jeffrey 1991; Stalnaker and Jeffrey 1994). McGee (1989) provides an axiomatic framework for compound conditionals: the rules (C1)–(C8) (McGee 1989, 504) are axioms for a probability calculus for conditionals. The corresponding semantics is based on the idea of the selection function on possible worlds given in Stalnaker (1968): for every world w and proposition A , $f(w, A)$ is the w -nearest world in which A is true. The probability of $A \rightarrow B$ is then computed by using the probability distribution of such selection functions f (and of course, we have to check what happens to proposition B in these w -nearest A -worlds).¹⁶ This leads to an interesting construction of the probability space, with important consequences (however, it is formally quite intricate). For instance, it is possible to prove that every probability distribution defined on atomic propositions extends in a unique way to right-nested conditionals of arbitrary length (Theorem 4, McGee 1989, 507).¹⁷ One of the features of McGee's account is that the Import-Export Principle is satisfied—indeed, it is one of the postulates.¹⁸ So, McGee's account (implicitly) presumes the deep interpretation of the conditional (see Sect. 2 for the preliminary discussion on deep and shallow interpretation, and Sect. 5 for the formal models), as the Import-Export Principle is a typical feature of it.

An interesting proposal was made by van Fraassen (1976) and developed in a series of papers by Kaufmann (2004, 2005, 2009, 2015). It is based on the idea of a Stalnaker Bernoulli probability space which consists of infinite sequences of possible worlds (not worlds simpliciter).¹⁹

This model has the advantage of being general; in particular it also makes it possible to compute not only the probabilities of right-nested conditionals $A \rightarrow (B \rightarrow C)$, but also of left-nested conditionals $(A \rightarrow B) \rightarrow C$, conjoined conditionals $(A \rightarrow B) \wedge (C \rightarrow D)$, and conditional conditionals $(A \rightarrow B) \rightarrow (C \rightarrow D)$.²⁰ However, there is also a price to pay: the Stalnaker Bernoulli space consists of all infinite sequences, which makes the computations tedious, and imposes certain limitations, especially

¹⁶ Bradley (2012) proposes modifying this account by considering vectors of possible worlds, not worlds.

¹⁷ In our model this is also true; however, the result holds not only for the deep interpretation, but also for the shallow one. Moreover, it is possible to compute probabilities of right-nested conditionals with “mixed” interpretations, see Sect. 7.

¹⁸ To be more precise, the axiom states that $P((A \rightarrow (B \rightarrow C)) \leftrightarrow ((A \wedge B) \rightarrow C)) = 1$ which means, that the probability of this equivalence is 1.

¹⁹ We give a short formal description of the Stalnaker Bernoulli model in Sect. 4.

²⁰ Kaufmann's ideas were an important source of inspiration for our model, and we are in an intellectual debt. In particular, the discussion concerning the possible interpretations of the conditional (in particular deep/shallow for the nested conditional) were of importance for us.

for modeling more complex situations and contexts.²¹ But even more importantly, the way the probability space is defined imposes the shallow interpretation of right-nested conditionals, which leads to unintuitive consequences in many cases. This approach is not compatible with McGee's approach, as the Import-Export Principle generally fails.²²

Kaufmann's formula for the probability of the right-nested conditional computed within the Stalnaker Bernoulli space is:

$P_{BS}^*(A \rightarrow (B \rightarrow C)) = P(BC \mid A) + P(B^c \mid A)P(C \mid B)$, where B^c is the complement of B (i.e. it corresponds to $\neg B$) and BC is the intersection of B and C .

P is the function in the initial probability space, where the sentences A, B, C are represented (as events A, B, C) but no events correspond to the conditionals in this initial space. P_{BS}^* is the probability distribution within the Stalnaker Bernoulli space, where conditionals are interpreted as events (and have a formally defined probability).

The advantage of this approach is that it provides an explicit formula for computing the probabilities, but its scope of applicability is limited. Consider the wet match example: *If the match is wet, then it will light if you strike it*, which has the form $W \rightarrow (S \rightarrow L)$. Assume that the probabilities are as follows (Kaufmann 2005, 206): (a) of the match getting wet = 0.1; (b) that you strike it = 0.5; (c) that it lights given that you strike it and it is dry = 0.9; (d) that it lights given that you strike it and it is wet = 0.1. Moreover, striking the match is independent of its wetness. Kaufmann's formula gives the result $P_{BS}^*(W \rightarrow (S \rightarrow L)) = 0.46$, which is counter-intuitive as we would expect the probability to be 0.1. Kaufmann proposes a modification of the Stalnaker Bernoulli model which incorporates causal dependencies (Kaufmann 2005, 2009). It is based on the idea of using random variables (following Stalnaker and Jeffrey 1994) and it gives the proper result, i.e. 0.1 for the wet match example. The formula for the probability of the conditional in this model is very simple:

$$P^*(A \rightarrow (B \rightarrow C)) = P(C \mid AB)$$

(i.e. it obeys the Import-Export Principle). So, both the interpretations (deep and shallow) can be accounted for by the known models, but for each of the interpretations a different model is needed. There is no way to account for both within **one** model just by "adjusting its parameters".

²¹ In order to define the probability space in a formal way, we have to define the σ -field. This is generated by a kind of cylindric subset of Ω^* , and the probability measure is an extension of the finite additive measure defined for these subsets. So, we need to invoke a theorem of measure theory which is not very elementary.

²² However, it is important to notice that both models give the same formula for the probability of conjoined conditionals, i.e. $(A \rightarrow B) \wedge (C \rightarrow D)$:

$$P((A \rightarrow B) \wedge (C \rightarrow D)) = \frac{P(ABCD) + P(A^cCD) \cdot P(B \mid A) + P(ABC^c)P(D \mid C)}{P(A \vee C)}$$

This formula can also be obtained within our model; we consider this compatibility to be an argument for the thesis that it is a good approach to the problem.

5 The graph model for a simplified case

5.1 The general idea

Our model for nested conditionals will be given in terms of Markov graphs and in this section we present all needed formal tools. To introduce the idea, we first present the model for a simple conditional of the form $A \rightarrow B$.²³ Imagine John who buys coffee every day on his way to work. He always buys Latte, Espresso or Cappuccino (denoted by L , E , C respectively). John is choosing coffee at random (but with fixed probabilities). The conditional in question is $\neg L \rightarrow E$:

If John does not order a Latte, he will order an Espresso.

How are we going to compute its probability? First, we need a formal representation of John's habits, which is given by a probability space $S = (\Omega, \Sigma, P)$, with $\Omega = \{L, E, C\}$; $P(E) = p$; $P(C) = q$; $P(L) = r$. The σ -field Σ is the power set 2^Ω (it consists of 8 events). But obviously—as both common sense and the triviality results show—there is no event in the space $S = (\Omega, \Sigma, P)$ corresponding to the conditional $\neg L \rightarrow E$.²⁴ This means that we have to define another probability space $S^* = (\Omega^*, \Sigma^*, P^*)$ where the conditional $\neg L \rightarrow E$ will be given an interpretation as an event (i.e. as a measurable subset of Ω^*). To avoid misunderstanding, we will use the symbol $[a]$ to denote the event in the constructed probability space corresponding to the proposition a . For instance, $[\neg L \rightarrow E]^*$ will be the event in the space $S^* = (\Omega^*, \Sigma^*, P^*)$ corresponding to the proposition *If John does not order Latte, he will order an Espresso*. Informally speaking, this new space S^* will consist of sequences of events (scenarios) which decide the conditional, and the probabilities of such sequences will be defined in a straightforward manner using probabilities from the sample space S . Here we present the general idea and motivations, and a formal definition of S^* for the general case is given in Sect. 5.4.

Of course, we have to assume that it is reasonable to speak of circumstances in which the status of the conditional is settled. The notion of truth conditions is essential in our account, in particular they will play a role in the construction of the probability space S^* . Indeed, we ascribe the probabilities to sentences in terms of fair bets: how much are we going to bet on the conditional a i.e. on winning the “ a - Conditional Game”?²⁵ In order to give a formal model, we need to know exactly

²³ In Wójtcowicz and Wójtcowicz (2019) we give a detailed study of possible interpretations of the conditional based on graph formalism, i.e. the global interpretation and the local interpretation (cf. Kaufmann 2004, 2005, 2009, 2015; Khoo 2015, 2016). In particular, we show that it is possible to defend (a suitable reformulation of) the P*CCP principle. We also hope that it will be possible to apply our account to Khoo's SP theory (2016), who obtained important results concerning the context sensitivity of (the probability of) conditionals. (However, the discussion exceeds the scope of the present study).

²⁴ In the probability space $S = (\Omega, \Sigma, P)$ there are eight events $\{\emptyset, \{C\}, \{E\}, \{L\}, \{C,E\}, \{C,L\}, \{E,L\}, \{C,E,L\}\}$. Sentences without the conditional operator, like *John ordered Latte* or *John did not order Cappuccino* have obvious interpretations in this space: $\{L\}$ and $\{L,E\}$ respectively. But which of these eight events should be considered to be the interpretation of the conditional *If John does not order Latte, he will order an Espresso*? Is it $\{E\}$? Or perhaps $\{E,C\}$? There are no reasonable candidates in $S = (\Omega, \Sigma, P)$, we need a different space.

²⁵ Acting on a belief (and betting is a classic example) is often considered to be a better measure of (subjective) probability than the value obtained by introspective analysis.

how the conditional α is interpreted, i.e. what the rules of the game are. In particular, we need to know when the bet is won, when it is lost, and how we should proceed if it is still undecided.

So, we can think of our toy example in terms of a “Non-Latte-Espresso Game”: every day John orders a coffee and evaluating the probability of $\neg L \rightarrow E$ depends on the possible scenarios. If it is an Espresso or Cappuccino, the game finishes and the bet is decided (Espresso—we win; Cappuccino—we lose). But it is of critical importance for the model to decide what to do when John orders Latte (i.e. when the antecedent is not fulfilled). *A priori* there are four possibilities:

1. We win the game.
2. We lose the game.
3. The game is cancelled (and nothing else happens).
4. The game is continued until we are able to settle the bet.

The first option reduces the conditional $\neg L \rightarrow E$ to a material implication, but this is not the right solution. The second choice is even less natural as it means we really played the “Neither-Latte-nor-Cappuccino Game”, without any conditional involved. The third choice is pessimistic: it means that we cannot model the probability of the conditional in cases when the antecedent is not fulfilled (so, in fact we abandon the construction of the model).²⁶ So, we choose the last solution which means that the bets are not settled and the game continues.²⁷ We assume that it is reasonable to think of performing a series of trials with unchanging conditions (“travelling across possible worlds”).²⁸ This idea is expressed by, for example, van Fraassen: “Imagine the possible worlds are balls in an urn, and someone selects a ball, replaces it, selects again and so forth. We say he is carrying out a series of Bernoulli trials” (van Fraassen 1976, 279). So, the truth conditions (which enable the probability of the conditional to be evaluated) are formulated in terms of sequences of possible worlds.

Of course, if we speak of restarting the game this means that the probabilities of the atomic events have not changed. These are some possible scenarios of the game:

²⁶ De Finetti discusses conditionals with a false antecedent proposing an additional logical value in his model (see for instance Douven 2016 for analysis). But this approach strongly differs from ours, as a third logical value is postulated. Speaking in terms of bets concerning $A \rightarrow B$, if the antecedent A is not realized the bet is called off (in our model we proceed with the game in the appropriate way). Importantly, in de Finetti’s approach right-nested conditionals obey the Import-Export Principle (which in this context means that the truth values of $A \rightarrow (B \rightarrow C)$ and $(A \wedge B) \rightarrow C$ coincide with respect to the de Finetti’s truth tables). But this means that the shallow interpretation is excluded (see Sect. 6.3.1.2), so de Finetti’s approach is more restrictive. (For discussion and formal results of the trivalent approach to conditionals see Égre et al. 2019, which focuses on the logical issues—however Adams’ thesis is also mentioned and it is formulated in terms of assertability). We thank the anonymous referee for helpful suggestions on this matter.

²⁷ McGee observes that “We imagine that, once she has set the price of a fair standard bet, the agent is required [...] to buy any number of standard bets that A if the price is $\$Pr(A)$ or less” (McGee 1989, 494); this means that the bet is repeated. (McGee also gives an analysis in terms of an insurance purchase, which might be more intuitive.).

²⁸ In the MED example, this means that we proceed with the observation of John’s behavior over the next day(s).

- E (Espresso)—we WIN.
- LE (Latte, Espresso)—we WIN.
- LL...LE(Latte, Latte, ..., Latte, Espresso)—we WIN.
- C (Cappuccino)—we LOSE.
- LC (Latte, Cappuccino)—we LOSE.
- LL...LC (Latte, Latte, ..., Latte, Cappuccino)—we LOSE.

It is clear that scenarios of this form exhaust the class of scenarios that settle the bet. They cannot be represented in the simple probability space $\Omega = \{L, E, C\}$ which models John's single choices, therefore we need a different space in which the scenarios (i.e. possible courses of the game) can be imbedded and their probabilities evaluated. The to-be-constructed probability space $S^* = (\Omega^*, \Sigma^*, P^*)$ will consist of scenarios which settle the game (the bet)—i.e. of the form $L...LE$ or $L...LC$ (these sequences will be elementary events i.e. elements of the Ω^*). We have to ascribe appropriate probabilities to such sequences, and we make use of the fact, that in the space $S = (\Omega, \Sigma, P)$ probabilities of choosing Espresso, Cappuccino or Latte are given (p, q, r) respectively). It is clear that the probability of ordering three Lattes in a row and a “follow-up Espresso” (i.e. $LLLE$) is $rrrp$ (so we set $P^*(LLLE) = r^3p$). Formal details are elaborated in Sect. 5.4.

We will use the formalism of Markov chains, which are an important tool in stochastic modeling in physics, biology, chemistry, finance analysis, social sciences etc.²⁹ However, they are not so widely used for modeling linguistic phenomena (however, see Bod et al. 2003), thus for the convenience of the reader we present the relevant notions here. Regardless of the technical details, the intuitive idea is fairly simple: we think about a system evolving in time (time being discrete) which is always in one of a range of possible internal states.³⁰ It is helpful to view the evolution of the system as a result of a series of actions occurring with fixed probabilities (like tossing a coin, choosing a coffee, drawing a ball from an urn etc.).

First, we give an illustration using a classic example of a Markov chain; the general definitions are given later.

5.2 Gambler's Ruin

Consider a coin flipping game, one of the player is the Gambler. At the beginning of the game, the Gambler has one penny, the opponent has two pennies.³¹ After each flip of the coin, the loser transfers one penny to the winner and the game restarts on the same terms. Assume that the Gambler bets on Heads; fix the probability of

²⁹ In the text we will instead use the term “Markov graphs” as the graphical interface is important to provide understanding.

³⁰ We consider Markov **chains**, i.e. processes with discrete time (time moments indexed by natural numbers). In the general case of Markov **processes**, time is modeled by (a subset of) the set of real numbers.

³¹ One versus two pennies is the simplest non-trivial case: if both have one penny, the game stops after the first toss.

tossing Heads to be p , and of tossing Tails to be q ($p + q = 1$). Formally this means that we have defined a simple probability space $S = (\Omega, \Sigma, P)$ describing a single coin toss with: $\Omega = \{H, T\}$; $\Sigma = 2^\Omega$; $P(H) = p$, $P(T) = q$. Obviously, the coin does not remember the history of the game so the probabilities (of tossing Heads/Tails) do not change throughout the whole game. This simple feature of Markov chains (i.e. memorylessness) is crucial for our purposes.

We can represent the current state of the game as the number of the Gambler's pennies, which means that the possible states of the system are 0, 1, 2, 3. The game starts in state 1 (one penny) and the transitions between states occur as a consequence of one of two actions: the coin coming up Heads or Tails (H,T). The game stops when it enters either state 0 (Gambler's Ruin) or state 3 (Gambler's victory). They are called **absorbing** states.

The game allows for a very natural graphical representation, with the arrows indicating the transitions between the states (p, q being the respective probabilities). For instance, the probability of getting from 1 to 2 is p ; the probability of getting from 1 to 0 is q . Both states 0 and 3 are absorbing—so with probability 1 the process terminates (formally: remains) in these states.

The graphical representation of the dynamics of the game is given in Fig. 1.

It is convenient to think of the process (for instance of the scenario of a game) in terms of travelling within the graph, from one vertex to another, until the process stops. The vertices represent the states of the system, the edges represent transitions and we can think of them as representing actions. Tossing H (Heads) means, that the Gambler wins one penny, so the state changes from n to $n + 1$ ($1 \mapsto 2$ or $2 \mapsto 3$). Tossing T (Tails) means, that the Gambler has to transfer one penny to his opponent, so the state changes from n to $n - 1$ ($2 \mapsto 1$ or $1 \mapsto 0$). So we can naturally track the possible scenarios of the game (i.e. ending either with victory or with loss) as possible paths in the graph.

Here are some examples of winning paths: HH (tossing Heads twice) generates the path $1 \mapsto 2 \mapsto 3$; HTHH generates $1 \mapsto 2 \mapsto 1 \mapsto 2 \mapsto 3$; HT...HTHH generates $1 \mapsto 2 \mapsto \dots \mapsto 1 \mapsto 2 \mapsto 1 \mapsto 2 \mapsto 3$.

Scenarios (paths) for losing the game are: T ($1 \mapsto 0$); HTT ($1 \mapsto 2 \mapsto 1 \mapsto 0$); HT...HTT ($1 \mapsto 2 \mapsto 1 \mapsto \dots \mapsto 2 \mapsto 1 \mapsto 0$).

It is clear what the probabilities of these scenarios are: for instance the probability of HTHH is $pqqp$ (the tosses of the coins are independent, so we simply multiply the probabilities).

Of course these sequences do not appear in the initial probability space $S = (\Omega, \Sigma, P)$. This means that we start with a simple probability space $S = (\Omega, \Sigma, P)$ and we

Fig. 1 Gambler's ruin

have to construct an appropriate “derived” probability space $S^* = (\Omega^*, \Sigma^*, P^*)$ (depending on the rules of the game). It is helpful to view the new space as “generated” by the graph as possible travel scenarios.

We denote this probability space by $S^* = (\Omega^*, \Sigma^*, P^*)$, the star $*$ indicates that it is different from the simple initial probability space S . The Ω^* consists of all sequences settling the Gambler’s Ruin game, i.e.:

$$\Omega^* = \{(HT)^n HH, (HT)^n T : n \geq 0\},$$

(here $(HT)^n$ means the sequence HT repeated n times, i.e. $(HT)^n = HT \dots HT$).

The probabilities of the elementary events are set according to the following rules:

$$\begin{aligned} P^*(HH) &= pp = p^2 \\ P^*(HTHH) &= pqpp = p^3 q \\ &\dots \\ P^*((HT)^n HH) &= (pq)^n pp = p^{n+2} q^n \end{aligned}$$

Similarly:

$$\begin{aligned} P^*(T) &= q \\ P^*(HTT) &= (pq)q = pq^2 \\ &\dots \\ P^*((HT)^n T) &= (pq)^n q = p^n q^{n+1} \end{aligned}$$

Computing the probability of the Gambler’s victory within $S^* = (\Omega^*, \Sigma^*, P^*)$ amounts to computing the probability of the event in S^* which is the counterpart (interpretation) of the sentence *The Gambler wins the game*. This event consists of all winning scenarios. In the paper we use $[\alpha]$ to denote the event corresponding to the sentence α in the appropriate probability space, so we can write:

$$[\textit{The Gambler wins the game}] = \{(HT)^n HH : n \geq 0\}$$

Therefore:

$$P^*([\textit{The Gambler wins the game}]) = P^* (\{(HT)^n HH : n \geq 0\}) = \sum_{n=0}^{\infty} (pq)^n p^2 = \frac{p^2}{1-pq}$$

The probability of the Gambler’s Ruin is $\frac{1-p}{1-pq}$. For $p = 1/2$ (a fair coin) the probability of the Gambler’s Ruin is $2/3$; the probability of winning is $1/3$.

The power of the Markov graph theory is revealed by the fact, that the corresponding probabilities can be computed in a direct way by examining the graph and solving a simple system of linear equations. Indeed, let $P(n)$ denote the probability of winning the Gambler’s Ruin game, which starts in state n , for $n = 1, 2$. Obviously:

$P(0) = 0$, as The Gambler has already lost;
 $P(3) = 1$, as The Gambler has already won.

To compute the probabilities $P(1)$, $P(2)$ we reason in an intuitive way. Being in the state n , we can either:

- toss Heads with probability p —then we are transferred to the state $n + 1$, and our chance of winning becomes $P(n + 1)$;
- toss Tails with probability q —then we are transferred to the state $n - 1$, and our chance of winning becomes $P(n - 1)$.

So, $P(1) = pP(2) + qP(0)$ and $P(2) = pP(3) + qP(1)$. Our system of equations is therefore:

$$\begin{aligned} P(0) &= 0 \\ P(1) &= pP(2) + qP(0) \\ P(2) &= pP(3) + qP(1) \\ P(3) &= 1 \end{aligned}$$

The solution (i.e. the probability of the Gambler's victory) is (as before)
 $P(1) = \frac{p^2}{1-pq}$.

An important feature of the gambling process is that the probabilities of atomic actions H and T remain fixed throughout the whole process (homogeneity), and the history of the game does not influence the effects (memorylessness). This means that the next step depends only on the present state of the process, not on the whole past. (This applies to our Non-Latte-Espresso example: we assume that the probabilities of choosing Latte, Espresso or Cappuccino are fixed—they do not depend on the type of coffee John ordered the previous time.)

In the Gambler's ruin game, we distinguish in a natural way three states: START (i.e. 1), WIN (i.e. 3) and LOSS (i.e. 0). This will be the standard situation when constructing graphs representing conditionals because (i) our Conditional Games always have to start and (ii) we need to be able to decide whether the game was won or lost.

For completeness we now present some formal definitions concerning Markov chains.

5.3 The general definition of a Markov chain

Markov chains are random processes with discrete time (formally: sequences of random variables, indexed by natural numbers),³² which satisfy the property of memorylessness, i.e. the Markov property. A Markov chain is specified by identifying:

³² A readable presentation of the Markov chains theory is chapter 7 of Bertsekas and Tsitsiklis (2008). The classic reference, with an extensive discussion of the motivations is Feller (1968); see also chapter 11 of Grinstead and Snell (1997) for more general results. For applications in linguistics see Bod et al. (2003).

1. A set of states $S = \{s_1, \dots, s_N\}$;
2. A set of **transition probabilities** $p_{ij} \geq 0$, for $i, j = 1, \dots, N$ —i.e. the probabilities of changing from state s_i to state s_j . The probabilities obey the equations: $\sum_{k=1}^N p_{ik} = 1$, for $i = 1, \dots, N$ (which means, that the probabilities of going from state s_i to any of the possible states sum up to one). These probabilities are fixed throughout the whole process.

Any Markov chain can be described by a N -by- N matrix, with the probabilities p_{ij} being the respective entries.³³ States with $p_{ii} = 1$ (which means, that the system will never get out of the state s_i) are called **absorbing**. In general there need not be any absorbing states (we might want to simulate a never-ending or a very long-term process), or there might be many of them. But in the article we restrict ourselves to the case where there are two absorbing states: WIN and LOSS; we also specify a state START (as in The Gambler's Ruin game). We also assume that there is a path leading from any state to one of the absorbing states.

5.4 The probability space $S^* = (\Omega^*, \Sigma^*, P^*)$ associated with the Markov graph

We want to use Markov graphs to model processes in which the system in question might change its state after some action/event have taken place. The structure of the Markov graph in question depends on the situation we want to model, for instance on the rules of the game, the causal nexus *etc.* The graph exhibits the dynamics of the system, allows to track its evolution and represents the probabilities of the system to pass from one state to another.

The evolution of the system depends on the actions, which affect the state of the system. An action is represented within the graph as a transition from one state to another, i.e. as an arrow in the graph. Of course, which of the actions are ascribed to the arrows in the graph depends on the situation we want to model. A clear example is provided by the Gambler's Ruin game.³⁴

Take $S = (\Omega, \Sigma, P)$ to be the initial probability space, in which the actions (tossing a coin, drawing a ball from an urn, buying coffee *etc.*) are represented. These actions can change the state of the system. The dynamics of the process is represented by the Markov graph G , and our aim is to construct the corresponding probability space $S^* = (\Omega^*, \Sigma^*, P^*)$, which represents the possible scenarios of the process (game). We are interested in modeling conditionals, so we restrict our

³³ For instance for our Gambler's Ruin game with states 0,1,2,3 the matrix is:

	0	1	2	3
0	1	0	0	0
1	q	0	p	0
2	0	q	0	p
3	0	0	0	1

³⁴ An arrow in the Markov graph is labeled by more than one action, if there are many actions in the system which cause a state transition. For instance, the Gambler might use a die with 1 playing the roles of Tails and 2,3,4,5,6 playing the role of Heads.

attention to graphs which have an initial state START and two absorbing states WIN and LOSE: when the system reaches one of the absorbing states, it comes to a stop. The general idea is very intuitive:

- (i) The set of elementary events Ω^* consists of sequences of actions (scenarios) which lead the system from the initial state START to an absorbing state (i.e. WIN or LOSS).
- (ii) The probability P^* of such a sequence is defined by multiplying the probabilities of the actions (which are defined in the initial space S).

Formally: let G be the Markov graph corresponding to the conditional in question, $S = (\Omega, \Sigma, P)$ be the sample probability space. The set of elementary events is $\Omega = \{A_1, \dots, A_m\}$, and the probabilities of elementary events in S are $P(A_i) = p_i$.³⁵ The Ω^* consists of finite sequences of events from Ω according to the following rule: the sequence $A_{i_1} \dots A_{i_n}$ is an elementary event in Ω^* if there is a sequence of states $s_{i_1} \dots s_{i_{n+1}}$ of the Markov graph G , such that:

- (i) $s_{i_1} = \text{START}$;
- (ii) $s_{i_{n+1}}$ is an absorbing state (i.e. WIN or LOSS) (and no of the states $s_{i_1} \dots s_{i_n}$ is an absorbing state);
- (iii) For any $k = 1, \dots, n$, A_{i_k} leads from the state s_{i_k} to $s_{i_{k+1}}$.

The probability P^* of the elementary event (i.e. sequence) $A_{i_1} \dots A_{i_n}$ is defined as $p_{i_1} \dots p_{i_n}$ (where $p_{i_k} = P(A_{i_k})$).³⁶ By definition, we consider finite paths only (the process has to terminate). Ω^* is therefore at most countable.

The respective σ -field Σ^* is the power set of Ω^* , i.e. $\Sigma^* = 2^{\Omega^*}$.

P^* has been defined for elementary events and it extends in a unique way to all sets $X \subseteq \Omega^*$ by the standard formula $P^*(X) = \sum_{\omega \in X} P^*(\omega)$.

$S^* = (\Omega^*, \Sigma^*, P^*)$ is the probability space corresponding to the graph. It should be stressed, that we cannot construct the graph knowing S only—we also have to know, how the events from S affect the game.³⁷ Therefore, the “ingredients” needed to define S^* are: (i) the initial probability space $S = (\Omega, \Sigma, P)$ and (ii) the Markov graph G which models the game. If we model a conditional α then the corresponding graph G_α has to represent its interpretation and the probability space S^*_α is based on G_α (and obviously the sample space S). This is clearly shown in the case of deep and shallow interpretations of the right-nested conditional (Sects. 5.1 and 5.2).

³⁵ The space $S = (\Omega, \Sigma, P)$ for the Gambler’s Ruin example consisted of two elementary events H, T with probabilities $P(H) = p$; $P(T) = q$. For the conditional *If it is even, then it is a six* the $\Omega = \{1, 2, 3, 4, 5, 6\}$ with all probabilities set to $1/6$ (assuming the die is fair).

³⁶ For instance, HTHH is one of the paths for the Gambler’s Ruin example (the corresponding set of states being 12123). Its probability is given by $P^*(HTHH) = pqpp$. The Ω^* for the Gambler’s Ruin game is $\Omega^* = \{(HT)^n HH, (HT)^n T: n \geq 0\}$, the probability P^* of each of these sequences is given by the obvious rule.

³⁷ For instance, it is not enough to know the probabilities of tossing Heads, Tails—it is necessary to know, how these action affect the system. We might stipulate different rules of such a kind of game (for instance: the second H is ignored), and in this case the graph G would be different. A die might be used for diverse games, and obviously the corresponding graphs have to represent the rules of the game.

5.5 Absorption probabilities of the Markov chain

We are interested in the probability of winning (losing) the game, i.e. in the absorption probabilities of the Markov chain. The idea has been given for the Gambler's ruin example. Now consider the general case of a Markov chain with N states $s_1, \dots, s_N$. To make the presentation consistent with the general approach in this text we assume that there are two absorbing states, say s_I (LOSS) and s_N (WIN).

The process is described by the transition probabilities p_{ij} , for $i, j = 1, 2, \dots, N$. As s_I and s_N are absorbing states, it means that:

$$\begin{aligned} p_{11} &= 1; & p_{1i} &= 0 & \text{for } i \neq 1. \\ p_{NN} &= 1; & p_{Ni} &= 0 & \text{for } i \neq N. \end{aligned}$$

Let $P(i)$ be the probability of eventually reaching state s_N , starting from state i . By definition, $P(1) = 0$ and $P(N) = 1$ (the game is already over).

Fact: The probabilities $P(i)$ are the unique solution of the system of equations:

$$\begin{aligned} P(1) &= 0 \\ P(N) &= 1 \\ P(i) &= \sum_{k=1}^N p_{ik} \cdot P(k), \quad \text{for } i = 2, \dots, N-1 \end{aligned}$$

So, we have in general a linear system of N equations with N variables, and the solution is straightforward.³⁸

We finish with the observation that (apart from some degenerate cases) the probability that the process will be eventually absorbed is 1. This means, that the probabilities of the finite sequences (i.e. the Ω^* in the constructed probability space) sum up to 1. This means, that we do not need any infinite sequences (paths) in our construction of $S^* = (\Omega^*, \Sigma^*, P^*)$.

5.6 The Markov graph for $\neg L \rightarrow E$

The game for the conditional *If John does not order a Latte, he will order an Espresso* is represented by the graph in Fig. 2.

Computing the probability of victory is simple. We take the corresponding probabilities of choosing Espresso, Cappuccino and Latte to be p , q , r . This means, that in our simple probability space $S = (\Omega, \Sigma, P)$ modeling John's choices, $P(E) = p$, $P(C) = q$, $P(L) = r$. The system of equations for this Markov graph consists of one equation only:

³⁸ Consider the Gambler's Ruin game. The respective system of equations is:

$$\begin{aligned} P(0) &= 0; \\ P(1) &= pP(2) + qP(0); \\ P(2) &= pP(3) + qP(1); \\ P(3) &= 1. \end{aligned}$$

Fig. 2 The graph for the simple conditional $\neg L \rightarrow E$

$$P_{\text{START}} = p + r \cdot P_{\text{START}}$$

P_{START} is the probability of getting from the initial point (i.e. START) to the state WIN. The result is $P_{\text{START}} = \frac{p}{1-r} = \frac{p}{p+q}$, which is equal to the conditional probability $P(E| \neg L)$ in the initial space.³⁹

The probability space $S^* = (\Omega^*, \Sigma^*, P^*)$ that represents the game is quite simple. Ω^* consists of all the possible scenarios (settling the bet), i.e. $\Omega^* = \{L^n E, L^n C : n \geq 0\}$. They might be viewed as paths in the graph, starting at START and ending at one of the absorbing states (either WIN or LOSS). The probabilities of the elementary events are:

$$\begin{aligned} P^*(L^n E) &= r^n p, \quad \text{for } n \geq 0; \\ P^*(L^n C) &= r^n q, \quad \text{for } n \geq 0. \end{aligned}$$

The σ -field Σ^* is the power set of Ω^* (as Ω^* is countable): $\Sigma^* = 2^{\Omega^*}$. The probability of any event $X \in \Sigma^*$ is defined in the standard way as the sum of the series

$$P^*(X) = \sum_{\omega \in X} P^*(\omega)$$

The interpretation of the conditional $\neg L \rightarrow E$ as an event in this space consists of the scenarios leading to a win, i.e.

$$[\neg L \rightarrow E]^* = \{L^n E : n \geq 0\}.$$

Its probability is:

$$P^*([\neg L \rightarrow E]^*) = \sum_{n=0}^{\infty} p r^n = \frac{p}{1-r} = \frac{p}{p+q}$$

³⁹ Even without the general theory, we can justify the equation in an intuitive way. What are our chances P_{START} of winning the game? There are two ways of winning:

- i. With probability p we can win in just one move.
- ii. With probability r , we will remain in the starting position. But then the game is restarted, so the probability of winning is again P_{START} .

The “contributions” of these variants to P_{START} are (i) p ; (ii) rP_{START} , which gives the equation $P_{\text{START}} = p + rP_{\text{START}}$.

Of course, this probability equals P_{START} . The probability space $S^* = (\Omega^*, \Sigma^*, P^*)$ corresponds to the graph; indeed, this is the probability space underlying Markov graph formalism (reassuring us that the equations for the graph are legitimate).

$$P^*([\neg L \rightarrow E]^*) = P_{\text{START}} = \frac{P}{1-r} = \frac{P}{p+q} = P(E|\neg L)$$

where P is the probability function from the initial probability space $S = (\Omega, \Sigma, P)$.

The event $[\neg L \rightarrow E]^*$ within $S^* = (\Omega^*, \Sigma^*, P^*)$ represents the conditional *If John does not order a Latte, he will order an Espresso*. But also the sentences *John buys an Espresso* (i.e. E) and *John buys a Cappuccino* (i.e. C) are interpreted in S^* in an obvious way: $[E]^* = \{E\}$; $[C]^* = \{C\}$. This is possible as these one-element sequences describe the simplest scenarios settling the bet. Also the sentence *John buys a Latte* is given an interpretation in $S^* = (\Omega^*, \Sigma^*, P^*)$, but here we have to be careful. As our aim is to represent the **settled** games, so we need not (and even cannot) represent in our space the claim *John buys a Latte—and this is the end of the whole story* (this would require the scenario “ L ” which is not present in S^*). But the interpretation of *John buys a Latte* consists of the set of sequences starting with L . Formally, $[L]^* = \{L^n E, L^n C : n \geq 1\}$. Observe, that $P^*([E]^*) = P(E)$; $P^*([C]^*) = P(C)$ and $P^*([L]^*) = P(L)$, which means that the probabilities of events from S have been preserved in S^* (which is a natural and desirable situation).⁴⁰

The construction of the graph and probability space for simple conditionals $A \rightarrow B$ is well behaved with respect to the negation. We expect the probability of $\neg(\neg L \rightarrow E)$, i.e. *It is not the case that if John does not order a Latte, he will order an Espresso* to be $1 - P^*([\neg L \rightarrow E]^*)$. In our case losing the game *If John does not order a Latte, he will order an Espresso* amounts to winning the game *If John does not order a Latte, he will order a Cappuccino*. We can use our space $S^* = (\Omega^*, \Sigma^*, P^*)$ to compute its probability, as the sentence $\neg L \rightarrow C$ has an interpretation $[\neg L \rightarrow C]^* = \{L^n C : n \geq 0\}$. Finally, $P^*([\neg(\neg L \rightarrow E)]^*) = P^*([\neg L \rightarrow C]^*) = 1 - P^*([\neg L \rightarrow E]^*)$. This is obvious even without computing: all the scenarios which do not lead to victory lead to defeat.⁴¹ So, in general, $P^*([\neg(A \rightarrow B)]^*) = P^*([A \rightarrow \neg B]^*) = 1 - P^*([(A \rightarrow B)]^*)$.

5.7 The general features of the Stalnaker Bernoulli model

The Stalnaker Bernoulli models and Kaufmann’s results (see Kaufmann 2004, 2005, 2009; van Fraassen 1976) were a very important source of inspiration for our model. For the sake of comparison we consider the Stalnaker Bernoulli space built for the same set of atomic actions as in our example, i.e. $\{L, E, C\}$, and

⁴⁰ Here we encounter the general problem of combining non-conditional sentences with conditional sentences (and giving the appropriate probabilistic representation)—for instance we might think of sentences like $L \wedge (\neg L \rightarrow E)$. The example suggests a route, however the general problem exceeds the scope of the paper.

⁴¹ In general, this is reflected in the Markov graphs we construct by the fact that there are exactly two absorbing states: one of them is interpreted as a victory, and the other as a loss. So losing $A \rightarrow B$ is winning the game $\neg(A \rightarrow B)$ which amounts to winning $A \rightarrow \neg B$. This is coherent with Stalnaker (1968).

the conditional $\neg L \rightarrow E$. The corresponding Stalnaker Bernoulli space $S^{**} = (\Omega^{**}, \Sigma^{**}, P^{**})$ is defined in the following way:

- Ω^{**} is the set of all infinite sequences consisting of the events L, E, C;
- Σ^{**} is the σ -field defined over Ω^{**} ;
- P^{**} is a probability measure defined on Σ^{**} .

Informally we can think of the infinite sequences as unfolding scenarios (relative to which the conditional will be settled and its probability will be computed). Any particular infinite sequence of events (which come from the initial $\Omega = \{L, E, C\}$) has null probability. But the relevant events in the big space S^{**} modeling the conditionals are appropriate “bunches” of sequences—and these events have non-zero probabilities. This involves constructing an appropriate σ -field. It is not the power set of Ω^{**} (this would be too big), but the formal details of the construction need not concern us here.

In the Stalnaker Bernoulli space, the “bunches” are constructed from sequences starting with an initial finite (perhaps empty) sequence of L’s, followed by a E or C—but later an arbitrary infinite sequence follows. The probabilities of events modeling the appropriate scenarios of the games are as we expect them to be:

$$P^{**}(\{L^n EX^\infty : X^\infty \text{ is an arbitrary infinite sequence from L, E, C}\}) = r^n p.$$

$$P^{**}(\{L^n CX^\infty : X^\infty \text{ is an arbitrary infinite sequence from L, E, C}\}) = r^n q.$$

Such sets generate the appropriate σ -field. Within this space the appropriate construction can be conducted which leads to the result that:

$$P^{**}([\neg L \rightarrow E]^{**}) = \frac{p}{p+q} = P(E|\neg L)$$

So our result is consistent with the result within the Stalnaker Bernoulli model (and also to McGee’s account). In general, in all these models it is true that $P^*(A \rightarrow B) = P(B|A)$, which means that the P^* CCP principle holds. We use the star * to stress that P^* is the probability distribution in the (constructed) space S^* , whereas P is the probability distribution from the initial space S . A and B are sentences not containing the conditional operator.⁴²

6 The graph model for right-nested conditionals

We illustrate the general approach by a familiar example of drawing balls from an urn, which allows the crucial feature of the model to be exhibited (in particular, the possible interpretations of the compound conditional). The advantage of this illustration is that it fits very nicely with the possible worlds account, where a proposition is identified with a set of possible worlds (the balls might be thought of

⁴² In Wójtowicz and Wójtowicz (2019) we discuss different variants of the PCCP principle, in particular under the local and global interpretation of the conditional; we cannot discuss this issue in more detail in the present study.

as possible worlds which are truth makers). For the sake of the example, we assume that there are white, red and green balls in the urn, which means that the propositions “*The ball is White*”, “*The ball is Green*”, “*The ball is Red*” are identified with sets of balls. Of course, no event in this space corresponds to the conditional $\neg W \rightarrow G$ (*If the ball is not White, it is Green*) as no particular set of balls could be considered the right interpretation. Even without the triviality results, this is a very awkward idea.⁴³ In our approach, which is more of a “possible scenarios model” or a game-theoretic semantics, no problems of this kind arise.

Assume moreover that the balls have one other feature, for instance mass. So the balls are either Heavy or Light. We assume that these two “dimensions” (mass/color) are independent, so all possible sorts of balls can appear in the urn. We want to consider right-nested conditionals like:

$$H \rightarrow (\neg W \rightarrow G)$$

i.e. *If the ball is Heavy, then if it is not White, it is Green.*⁴⁴

As we think of this conditional in terms of a game (bet), we have to make sure what the exact terms of the game are.

There are six types of balls in our urn:

HW—Heavy White balls
 HG—Heavy Green balls
 HR—Heavy Red balls
 LW—Light White balls
 LG—Light Green balls
 LR—Light Red balls

Some events are of particular interest: $H = HW \cup HG \cup HR$; $L = LW \cup LG \cup LR$; $W = HW \cup LW$ etc.⁴⁵

There is a probability distribution (as there is some definite number of balls of each kind), so there is an initial probability space $S = (\Omega, \Sigma, P)$, with $\Omega = \{HG, HR, HW, LG, LR, LW\}$. In order to define the probability of the conditional $H \rightarrow (\neg W \rightarrow G)$, we need a probability space $S^* = (\Omega^*, \Sigma^*, P^*)$ where the conditional $H \rightarrow (\neg W \rightarrow G)$ is given an interpretation as an event $[H \rightarrow (\neg W \rightarrow G)]^* \in \Sigma^*$. As in the Non-Latte-Espresso Game, the probability space will be associated with the

⁴³ In the simplest case, there are three balls: one is White, one is Green, one is Red. Which set of balls is the event [*If the ball is not White, it is Green*]? There are 8 possibilities, and none of them is acceptable.

⁴⁴ In some cases, the properties of the negation might cause additional problems, but here the situation is clear enough. However, if we want to avoid the negation completely, we can simply use the term “Dark” for “non-White”, therefore our example becomes *If the ball is Heavy, then if it is Dark, it is Green*. We can reformulate our Espresso-Cappuccino-Latte example from Sect. 5.1 in the same way.

⁴⁵ If we are very meticulous about the notation, we should be very careful in distinguishing $\vee$ and $\cup$: $\vee$ applies to propositions, whereas $\cup$ applies to events that are interpretations of these propositions. In some cases, we will have to be careful to distinguish the proposition (in particular the conditional) from the event in the appropriate probability space as both the space and the event depend on the interpretation of the conditional. But if it does not lead to misunderstandings, we will switch between them by (a slight) abuse of language.

respective graph defining the rules of the game (depending on the interpretation of the conditional).

But before we start The Nested Conditional Game, we have to decide when the game is (1) won, (2) lost, or (3) continued. Two cases are obvious:

- If we draw an HG, i.e. a Heavy Green ball, we WIN;
- If we draw an HR, i.e. a Heavy Red ball, we LOSE.

But what happens in the other cases? The case of Light balls is not controversial:

- If it is a Light ball (of any color, i.e. an LW or LG or LR), we restart the game. Restarting means we put the ball back into the urn and draw again, of course still paying attention to the weight and color.⁴⁶

The last case remains: a Heavy White ball. Depending on the interpretation (deep versus shallow), the game is either:

1. restarted on the same terms;
2. continued on modified terms.⁴⁷

Both the interpretations lead to different graphs, different corresponding probability spaces, and different formulas for the probabilities.

6.1 The deep interpretation

The rules of the game for the deeply interpreted conditional are simple: we draw the first ball and if we draw (HG) or (HR), the game is decided. In all other cases, the game is not decided and restarts from the beginning, which means that we still pay attention to the weight! This is because if we interpret the conditional *If the ball is Heavy, then if it is not-White, it is Green* as if the ball is Heavy, then **this particular ball** has to fulfill the condition: *if it is not-White, it is Green*. The graph for the game is given in Fig. 3:

⁴⁶ In Wójtowicz and Wójtowicz (in preparation) we analyze the situation in which the ball is **not** put back, so that every turn of the game changes the probability distributions. This represents the fact that habits and the respective probabilities can change, and it also has a natural interpretation in terms of possible worlds' histories.

⁴⁷ Consider the MED example—and the day, when John took A but did not take B. Depending on the interpretation it either means that the game: (1) is restarted, i.e. we wait for the day when John takes both A and B (deep interpretation), or that (2) it continues on different terms, i.e. we only have to wait until John takes B.

Fig. 3 The graph for the deep interpretation of the conditional $H \rightarrow (\neg W \rightarrow G)$

There is only one equation to solve:

$$P_{\text{START}} = P(\text{HG}) + rP_{\text{START}}$$

where $r = P(\text{HW}) + P(\text{LW}) + P(\text{LG}) + P(\text{LR})$

The solution is straightforward: $P_{\text{START}} = \frac{P(\text{HG})}{P(\text{HG}) + P(\text{HR})}$

We need to define a probability space $S_{\text{deep}}^* = (\Omega_{\text{deep}}^*, \Sigma_{\text{deep}}^*, P_{\text{deep}}^*)$ in which the conditional $H \rightarrow (\neg W \rightarrow G)$ has a counterpart as an event. The space is connected with the graph in an obvious way by considering the paths within the graph which lead to victory: they consist of some loops (“dummy moves”) followed by a decisive move. The “dummy moves” are: HW, LW, LG, LR (nothing happens, the game is restarted). The decisive moves are HR and HG. So, any such path will have the form $X^n(\text{HG})$, $X^n(\text{HR})$, where X is one of HW, LW, LG, LR. A winning path is, for example, (HW)(LR)(LG)(HG), a losing path is, for example, (HW)(HW)(LG)(LG)(LR)(HR). So:

$$\Omega_{\text{deep}}^* = \{X^n(\text{HG}), X^n(\text{HR}): n \geq 0\}$$

The probability of any particular path is computed by multiplying the probabilities. For instance, the probability of the path (LG)(LG)(LR)(HW)(LR)(HG) is $P(\text{LG}) \cdot P(\text{LG}) \cdot P(\text{LR}) \cdot P(\text{HW}) \cdot P(\text{LR}) \cdot P(\text{HG})$. The set of winning scenarios consists of admissible scenarios ending with a HG (Heavy Green) ball, i.e.

$$[H \rightarrow (\neg W \rightarrow G)]_{\text{deep}}^* = \{X^n(\text{HG}): n \geq 0\}$$

Its probability is easily computed within the space S_{deep}^* :

$$P_{\text{deep}}^*(H \rightarrow (\neg W \rightarrow G)) = \sum_{n=0}^{\infty} P(\text{HG}) \cdot (P(X))^n = \frac{P(\text{HG})}{1 - P(X)} = \frac{P(\text{HG})}{P(\text{HG}) + P(\text{HR})}$$

Of course, this agrees with the result computed within the graph, i.e.

$$P_{\text{START}} = \frac{P(\text{HG})}{P(\text{HG}) + P(\text{HR})}$$

To sum up: $P_{\text{deep}}^*(H \rightarrow (\neg W \rightarrow G)) = P_{\text{START}} = \frac{P(\text{HG})}{P(\text{HG}) + P(\text{HR})}$

This is analogous to Alice’s probability for MED (Sect. 2). Applying this formula to the wet match probabilities from Sect. 3, we get:

Fig. 4 The graph for the deep interpretation of the conditional $A \rightarrow (B \rightarrow C)$

$$P_{\text{deep}}^*(W \rightarrow (S \rightarrow L)) = 0.1$$

which agrees with Kaufmann's result obtained within the causal model, (Kaufmann 2005, 209–210).

Informally, we might say that under the deep interpretation of the conditional $H \rightarrow (\neg W \rightarrow G)$ we really need only observe Heavy balls, as only among these balls can the Nested Conditional Game be settled. In a sense, our game is reduced to the Colorful Conditional Game played with Heavy balls only.⁴⁸ Light balls might be ignored as they do not matter (which means the color distribution among Light balls does not matter).⁴⁹

For completeness, we conclude with presenting the deep interpretation in the general case $A \rightarrow (B \rightarrow C)$. The corresponding Markov graph is given in Fig. 4.

The probability space is defined in the following way (see the formal definition in Sect. 5.4):⁵⁰

$$\Omega_{\text{deep}}^* = \{((AB)^c)^n(ABC), ((AB)^c)^n(ABC^c) : n \geq 0\}.$$

The probabilities P_{deep}^* of the elementary events are defined in the obvious way:

$$\begin{aligned} P_{\text{deep}}^*(((AB)^c)^n(ABC)) &= (P((AB)^c))^n P(ABC) \\ P_{\text{deep}}^*(((AB)^c)^n(ABC^c)) &= (P((AB)^c))^n P(ABC^c) \end{aligned}$$

The interpretation of the conditional $A \rightarrow (B \rightarrow C)$ in S_{deep}^* is:

$$[A \rightarrow (B \rightarrow C)]_{\text{deep}}^* = \{((AB)^c)^n(ABC) : n \geq 0\}.$$

So, we have:

⁴⁸ We mean the $\neg W \rightarrow G$ game, but played with Heavy balls only. We can imagine a new urn, from which all the non-Heavy balls have been removed, and we pay attention to the colors only. From the mathematical point of view, we need a new probability space.

⁴⁹ “Skipping Light balls” means that Light balls are neutral “dummy” moves—like the days when John does not take medication A in MED for the deep interpretation (Sect. 3).

⁵⁰ In the text, we use the symbol X^c to denote the complement of set (event) X , so $(AB)^c$ is the complement of event AB . $((AB)^c)^n$ is the sequence consisting of n events $(AB)^c$.

$$P_{\text{deep}}^*(A \rightarrow (B \rightarrow C)) = \sum_{n=0}^{\infty} P(ABC)(P((AB)^c))^n = \frac{P(ABC)}{1 - P((AB)^c)} = \frac{P(ABC)}{P(AB)} = P(C|AB).$$

6.2 The shallow interpretation

Under the shallow interpretation of the conditional, after drawing the first Heavy White ball, the rules of the game change and the game continues in a different way.⁵¹ From now on we no longer pay attention to the weight, we only take colors into account. We might say informally that we lost the ability to distinguish weight, so we switched to the Purely Colorful Conditional Game ($\neg W \rightarrow G$)—still played within the same urn.⁵²

So, the interpretation of the conditional

If the ball is Heavy, then if it is not White, it is Green

can be described in the following way:

If I draw a Heavy ball, then—provided we continue the game (i.e. it has not been settled in this first move)—anytime later on if the ball is not-White, it is Green (regardless of its weight).

We might describe drawing a Heavy White ball as a kind of “Activating Event” which changes the conditions of the latter phase of the game, or we might think of it in terms of an updating rule which changes the terms of the game (or “transfers” us to a different game from now on).⁵³

The stochastic graph for the shallow interpretation is given in Fig. 5:

We can write down the equations for success in this graph:

$$\begin{aligned} P_{\text{START}} &= P(\text{HG})P_{\text{WIN}} + P(\text{L})P_{\text{START}} + P(\text{HW})P_{\text{COLORS}} + P(\text{HR})P_{\text{LOSS}} \\ P_{\text{COLORS}} &= P(\text{G}) + P(\text{W})P_{\text{COLORS}} \end{aligned}$$

P_{START} is the probability of winning the game when we start from START. P_{COLORS} is the probability of winning the game when starting from the node COLORS, which is the starting point of the Purely Colorful Conditional Game. $P_{\text{WIN}} = 1$, $P_{\text{LOSS}} = 0$ (the game is already finished). The solution is:

⁵¹ Remember unfortunate John who broke his leg in Davos (Sect. 3). The terms of his “life-game” have changed: it does not matter whether he goes skiing in Davos or anywhere else—his leg is more likely to hurt than before the accident.

⁵² Here the situation differs from the deep case: we do not remove any balls from the urn, but stop paying attention to their weight and distinguish their colors only.

⁵³ Applying it to the MED example, this interpretation is as follows: after the “Activation by taking medication A” we pay no longer attention to whether A has been taken on some day, but only whether taking B is accompanied by an allergic reaction.

Fig. 5 The graph for the shallow interpretation of the conditional $H \rightarrow (\neg W \rightarrow G)$

$$P_{\text{START}} = \frac{P(\text{HG}) + P(\text{HW}) \frac{P(\text{G})}{1 - P(\text{W})}}{P(\text{H})}$$

Which simplifies to:

$$P_{\text{START}} = P(\text{G}|\text{H}) + P(\text{W}|\text{H})P(\text{G}|\text{W}^c).$$

Just as in the deep case, we need to define a probability space $S_{\text{shallow}}^* = (\Omega_{\text{shallow}}^*, \Sigma_{\text{shallow}}^*, P_{\text{shallow}}^*)$ in which the conditional $H \rightarrow (\neg W \rightarrow G)$ has a counterpart as an event. And again, the space is connected to the graph in an obvious way and consists of these paths within the graph which lead to victory.

For completeness, we present the general construction for a conditional $A \rightarrow (B \rightarrow C)$. The Markov graph is analogous to the graph for $H \rightarrow (\neg W \rightarrow G)$ and has the form given in Fig. 6 (next page).

The space $S_{\text{shallow}}^* = (\Omega_{\text{shallow}}^*, \Sigma_{\text{shallow}}^*, P_{\text{shallow}}^*)$ consists of paths starting in START and terminating in one of the states WIN, LOSS. It is defined according to the formal definition from Sect. 4.4. This means, that:

$$\Omega_{\text{shallow}}^* = \{(A^c)^n(ABC), (A^c)^n(ABC^c), (A^c)^n(AB^c)(B^c)^k(BC), (A^c)^n(AB^c)(B^c)^k(BC^c) : n, k \geq 0\}.$$

For instance, $(A^c)^n(ABC)$ is the sequence consisting of n events A^c followed by the event ABC . $(A^c)^n(AB^c)(B^c)^k(BC)$ is the sequence consisting of n events A^c followed by the event AB^c , then by k events B^c and finally by the event ABC . Such sequences are elementary events in $\Omega_{\text{shallow}}^*$. Their probabilities P_{shallow}^* are obtained by multiplying the probabilities taken from the sample space $S = (\Omega, \Sigma, P)$. $\Sigma_{\text{shallow}}^*$ is the power set of $\Omega_{\text{shallow}}^*$, and P_{shallow}^* is defined by the familiar formula $P^*(X) = \sum_{\omega \in X} P^*(\omega)$. The structure of $\Omega_{\text{shallow}}^*$ is more complex than in the deep case. For

Fig. 6 The graph for the shallow interpretation of the conditional $A \rightarrow (B \rightarrow C)$

instance, there is the path $(AB^c)(A^cBC)$ which leads to victory, and the path $(AB^c)(A^cBC^c)$ which leads to loss (these paths do not even appear in Ω_{deep}^*).

The interpretation of the conditional $A \rightarrow (B \rightarrow C)$ in S_{shallow}^* is the event

$$[A \rightarrow (B \rightarrow C)]^* = \{(A^c)^n(ABC), (A^c)^n(AB^c)(B^c)^k(BC) : n, k \geq 0\}.$$

and its probability can be computed either by solving the system of equations for the Markov graph, or by direct computation in the space S_{shallow}^* (which involves some computation with infinite series). Regardless of the method, we get the formula:

$$P_{\text{shallow}}^*([A \rightarrow (B \rightarrow C)]^*) = P(BC|A) + P(B^c|A)P(C|B)$$

This agrees with Kaufmann's formula for the probability of a right-nested conditional under the shallow interpretation (in the Stalnaker Bernoulli model). Applying the formula to our example ($A = H$, $B = \neg W$; $C = G$), we get the already known result.

The shallow interpretation is natural in some cases, but examples where it leads to very unnatural consequences are also abundant, as the wet match example clearly shows.

Is it better to play the deep or the shallow game? It depends. In some cases, the probabilities coincide, sometimes P_{deep}^* is bigger, sometimes P_{shallow}^* . The formulas are:

$$P_{\text{deep}}^*(H \rightarrow (\neg W \rightarrow G)) = \frac{P(HG)}{P(HG) + P(HR)}$$

$$P_{\text{shallow}}^*(H \rightarrow (\neg W \rightarrow G)) = P(G|H) + P(W|H) \cdot P(G|W^c)$$

Consider three cases:

1. If $P(HW) = P(HG) = P(HR) = P(LW) = P(LG) = P(LR) = \frac{1}{6}$, then $P_{\text{deep}}^*(H \rightarrow (\neg W \rightarrow G)) = P_{\text{shallow}}^*(H \rightarrow (\neg W \rightarrow G)) = \frac{1}{2}$.
2. If $P(HW) = P(HG) = P(HR) = \frac{1}{6}$; $P(LW) = 0$; $P(LG) = 0$; $P(LR) = \frac{1}{2}$, then $P_{\text{deep}}^*(H \rightarrow (\neg W \rightarrow G)) = \frac{1}{2}$,
 $P_{\text{shallow}}^*(H \rightarrow (\neg W \rightarrow G)) = \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{5} = \frac{2}{5}$.
3. If $P(HW) = P(HR) = \frac{1}{6}$; $P(HG) = 0$; $P(LW) = 0$; $P(LG) = \frac{2}{3}$; $P(LR) = 0$, then $P_{\text{deep}}^*(H \rightarrow (\neg W \rightarrow G)) = 0$,
 $P_{\text{shallow}}^*(H \rightarrow (\neg W \rightarrow G)) = 0 + \frac{1}{2} \cdot \frac{4}{5} = \frac{2}{5}$.

6.3 Some specific issues

6.3.1 The Import-Export Principle

The Import-Export Principle states that:

$$P(A \rightarrow (B \rightarrow C)) = P((A \wedge B) \rightarrow C)$$

Is this true? We discuss this issue separately for the deep and the shallow interpretation. To keep things simple, we focus on the previously discussed example, i.e. on the conditional $H \rightarrow (\neg W \rightarrow G)$. The Import-Export Principle in this case would have the form:

$$P(H \rightarrow (\neg W \rightarrow G)) = P((H \wedge \neg W) \rightarrow G)$$

In both these formulas we have used a very simplified notation (and committed a slight abuse of language), as we really think about two different probability functions $P(\cdot)$ from different spaces. We have to be careful, but in this case it is clear what we mean.

6.3.1.1 The Import-Export Principle under the deep interpretation What is the probability of $((H \wedge \neg W) \rightarrow G)$? This is a simple conditional, the corresponding graph for which was defined in Sect. 6.1, and the corresponding probability is $\frac{P(HG)}{P(HG) + P(HR)}$, which agrees with $P_{\text{deep}}^*(H \rightarrow (\neg W \rightarrow G))$. This means that the Import-Export Principle holds for the deep interpretation. This is not surprising—intuitively, it should! We only pay attention to color **among** the Heavy balls, so we are interested in the outcome of the game only when the ball is Heavy **and** is non-White. Similarly, in the MED example (under the deep interpretation), we only observe John on the days when he has taken medication A.

So, for the deep interpretation the Import-Export Principle works, which means that we can compute the probabilities very easily.⁵⁴ The graph formalism allows to prove the following equation:

$$P_{\text{deep}}^*(A \rightarrow (B \rightarrow C)) = P^*((A \wedge B) \rightarrow C) = P(C|AB).$$

So, obviously, for the deep interpretation the antecedents of the conditional can be exchanged:

$$P_{\text{deep}}^*(A \rightarrow (B \rightarrow C)) = P_{\text{deep}}^*(B \rightarrow (A \rightarrow C)).$$

This is intuitive if we think of the Heavy/Green balls game: winning the *If not White, then Green* game played with the Heavy balls only is the same as winning the *If Heavy, then Green* game played with the non-White balls. This agrees with McGee's theory—the Import-Export Principle is his axiom (C7) (McGee 1989, 504).

6.3.1.2 The Import-Export Principle under the shallow interpretation If the Import-Export Principle was true for the shallow conditional, this would mean that:

$$P_{\text{shallow}}^*(A \rightarrow (B \rightarrow C)) = P^*((A \wedge B) \rightarrow C);$$

$$P_{\text{shallow}}^*(B \rightarrow (A \rightarrow C)) = P^*((B \wedge A) \rightarrow C).$$

Obviously, $A \wedge B \equiv B \wedge A$, so $P^*((A \wedge B) \rightarrow C) = P^*((B \wedge A) \rightarrow C)$. So, the IE Principle would imply that $P_{\text{shallow}}^*(A \rightarrow (B \rightarrow C)) = P_{\text{shallow}}^*(B \rightarrow (A \rightarrow C))$. But, in general

$$P_{\text{shallow}}^*(A \rightarrow (B \rightarrow C)) \neq P_{\text{shallow}}^*(B \rightarrow (A \rightarrow C)),$$

which means that for the shallow interpretation the Import-Export Principle does not hold. This is intuitive: the first condition changes the rules of the game, so depending on what the first condition is, the probabilities can differ greatly.⁵⁵ In particular, the order of the antecedents in the conditional is very important. The following graphs represent the two sentences $H \rightarrow (\neg W \rightarrow G)$ and $\neg W \rightarrow (H \rightarrow G)$, and they illustrate the general rule (Fig. 7):

⁵⁴ Using the formalism from Sect. 5.1. we can compute the probability of the conditional $P_{\text{deep}}^*(A \rightarrow (B \rightarrow C))$ as the probability of the corresponding event:

$$P_{\text{deep}}^*([A \rightarrow (B \rightarrow C)]_{\text{deep}}^*) = \sum_{n=0}^{\infty} P(ABC) \cdot P((AB)^c)^n = \frac{P(ABC)}{P(AB)} = P(C|AB)$$

In Sect. 5.6. we justified the equation $P^*(\neg L \rightarrow E) = P(E|\neg L)$ for special case of three events L, E, C. However, this can be easily generalized to $P^*([X \rightarrow Y]^*) = P(Y|X)$, for any events X, Y from the σ -field (as the particular features of L, E, C play no essential role). Take $X = A \wedge B$; $Y = C$ to obtain $P^*((A \wedge B) \rightarrow C) = P(C|A \wedge B)$.

⁵⁵ Consider MED: the problem whether A causes a permanent allergic disposition is in general very different from the problem, whether B causes such a disposition.

Fig. 7 The left graph corresponds to $H \rightarrow (\neg W \rightarrow G)$, the right graph to $\neg W \rightarrow (H \rightarrow G)$

6.3.2 P*CCP

From an abstract point of view, the P*CCP principle expresses the idea that it is possible to compute the probability of conditionals as a function of conditional probabilities in the sample space.⁵⁶ If we accept this general principle, then finding analogues of P*CCP amounts to expressing the probability $P^*(A \rightarrow (B \rightarrow C))$ in terms of conditional probabilities from the sample space S . We think that this retains some fundamental intuitions of Ramsey expressed in the well-known quotation (footnote 7).

Taken the problem in a more abstract setting, we want to know whether $P^*(A \rightarrow (B \rightarrow C))$ can be expressed as a function of conditional probabilities in the sample space S .⁵⁷ In other words, we want to identify the appropriate equation:

$$P^*(A \rightarrow (B \rightarrow C)) = \Psi(\dots \text{entries of the form of conditional probabilities within } S \dots).$$

For the simple conditional, $P^*(A \rightarrow B)$ is given via a particularly simple function $\Psi(\dots)$, namely $P^*(A \rightarrow B) = P(B|A)$. There is no obvious analogue for $A \rightarrow (B \rightarrow C)$ —for instance the equation $P^*(A \rightarrow (B \rightarrow C)) = P(B \rightarrow C|A)$ does not make sense, as $B \rightarrow C$ has no interpretation in the sample space S as an event. So, for right-nested conditionals $A \rightarrow (B \rightarrow C)$ the function $\Psi(\dots)$ in the analogue of P*CCP must be more complex. It is quite obvious, that the functions $\Psi(\dots)$ will be different for the deep and shallow interpretations, as $P_{\text{deep}}^*(A \rightarrow (B \rightarrow C))$ and $P_{\text{shallow}}^*(A \rightarrow (B \rightarrow C))$ differ.

6.3.2.1 P*CCP under the deep interpretation Under the deep interpretation, the Import-Export Principle holds, i.e.

$$P^*(A \rightarrow (B \rightarrow C)) = P^*((A \wedge B) \rightarrow C)$$

$(A \wedge B) \rightarrow C$ is a simple conditional, and its probability is given by the formula:

$$P^*((A \wedge B) \rightarrow C) = P(C|AB).$$

This means, that $P_{\text{deep}}^*(A \rightarrow (B \rightarrow C)) = P(C|AB)$. The function $\Psi(\dots)$ has therefore a very simple form, we really need only one conditional probability from S . It is also easy to show, that the following formula is true:

$$P_{\text{deep}}^*(A \rightarrow (B \rightarrow C)) = P_A(C|B)$$

P_A is the standard probability measure obtained from P by conditionalizing on A : $P_A(X) := P(X|A)$.⁵⁸ This means, that under the deep interpretation, the probability of

⁵⁶ We want to thank the anonymous referee for helpful observations concerning the presentation of P*CCP.

⁵⁷ I.e. we want to compute $P^*(A \rightarrow (B \rightarrow C))$ having at our disposal the values from the sample space: $P(A)$, $P(B)$, $P(C)$, $P(B|A)$, $P(C|A)$, $P(A|B)$, $P(C|B)$, $P(A|C)$, $P(B|C)$, $P(BC|A)$, $P(AC|B)$, $P(ABC)$, $P(B|AC)$, $P(C|AB)$, $P(A|BC)$ etc.

⁵⁸ $P(C|AB) = \frac{P(ABC)}{P(AB)} = \frac{P(ABC)}{P(A)} \frac{P(A)}{P(AB)} = \frac{P_A(BC)}{P_A(B)} = P_A(C|B)$.

the conditional $A \rightarrow (B \rightarrow C)$ equals probability of the conditional $B \rightarrow C$ relatively to a new sample probability space, obtained from S by conditionalizing on A . This is intuitive: consider our conditional $H \rightarrow (\neg W \rightarrow G)$. The probability of winning the “ $H \rightarrow (\neg W \rightarrow G)$ -game” is really the probability of winning the “ $(\neg W \rightarrow G)$ -game” restricted to Heavy balls only.⁵⁹

The equation $P_{\text{deep}}^*(A \rightarrow (B \rightarrow C)) = P_A(C|B)$ definitely looks like P^* CCP. And it satisfies the general requirement: probability of the (nested) conditional $A \rightarrow (B \rightarrow C)$ is expressed as a function of conditional probabilities from the sample space S , as the arguments of the function $\Psi(\dots)$ are $P(ABC)$ and $P(AB)$.

6.3.2.2 P^* CCP under the shallow interpretation For the shallow interpretation, the formula for the probability of the right-nested conditional is:

$$P_{\text{shallow}}^*(A \rightarrow (B \rightarrow C)) = P(BC|A) + P(B^c|A)P(C|B)$$

As in the deep case, the probability of the conditional $A \rightarrow (B \rightarrow C)$ is expressed as a combination of conditional probabilities from the initial sample space S . The function $\Psi(\dots)$ is more complex than in the deep case, but still fairly simple, with only three arguments: $P(BC|A)$, $P(B^c|A)$ and $P(C|B)$.⁶⁰

In conclusion, we can say that the P^* CCP principle holds for the deep interpretation in a straightforward way, the formula for equation $P_{\text{deep}}^*(A \rightarrow (B \rightarrow C))$ being formally quite analogous to the original formula $P^*(A \rightarrow B) = P(B|A)$. For the shallow interpretation there is no such purely formal analogy, but the probability $P_{\text{shallow}}^*(A \rightarrow (B \rightarrow C))$ is calculated as a function of conditional probabilities in the sample space S . We think that this retains the fundamental intuition of Ramsey (and PCCP).

7 Multiple-(right)-nested conditionals

We have considered conditionals of the form $A \rightarrow (B \rightarrow C)$, but we can also think of more complex conditionals $D \rightarrow (A \rightarrow (B \rightarrow C))$. Three ‘ $\rightarrow$ ’ connectives appear within this conditional. The distinction between the deep and shallow interpretation is not interesting for the simple conditional (either it does not make sense, or the interpretations coincide), so it cannot be reasonably applied to the last conditional. So, the ‘ $\rightarrow$ ’ between ‘ B ’ and ‘ C ’ is neutral in this respect, but the other two symbols (i.e. after A , and after D), can be given different interpretations. A priori there are four possible interpretations: deep-deep; deep-shallow; shallow-deep; shallow-

⁵⁹ Observe that not only the equation $P^*(A \rightarrow (B \rightarrow C)) = P_A(C|B)$ holds, but also $P^*(A \rightarrow (B \rightarrow C)) = P_B(C|A)$. This is a simple consequence of the fact, that under the deep interpretation the Import-Export Principle holds, so that A and B can be “switched” without changing the interpretation of the conditional.

⁶⁰ It is also possible to prove the equation $P_{\text{shallow}}^*(A \rightarrow (B \rightarrow C)) = P_{\text{shallow}}^*((B \rightarrow C)|A)$, which is formally more similar to PCCP. We think this is an interesting feature of the formal model. However, this equation does not provide a direct representation of $P_{\text{shallow}}^*(A \rightarrow (B \rightarrow C))$ as a combination of conditional probabilities in the initial sample space S , so we think that $P_{\text{shallow}}^*(A \rightarrow (B \rightarrow C)) = P(BC|A) + P(B^c|A)P(C|B)$ has more of the “ P^* CCP-spirit”.

shallow. In this section we consider only the uniform interpretations, i.e. deep-deep and shallow-shallow.

Consider the sentence:

If John takes medication D, then if he takes A, then if he takes B, then he will have an allergic reaction.

Depending on the interactions between the medications (some combinations might cause permanent changes, which would not occur if only one of the mediations was taken), different interpretations of the conditional will be justified.

Take the familiar example of Heavy/Light and White/Green/Red balls, but now we assume that still another “orthogonal” property can be defined, e.g. Big/Small.⁶¹ Now consider the conditional:

If the ball is Big, then if it is Heavy, then if it is not-White, it is Green

i.e. $B \rightarrow [H \rightarrow (\neg W \rightarrow G)]$. This conditional can be interpreted in different ways, and we can identify the interpretations in terms of graphs.

7.1 The deep-deep interpretation

If we assume the deep-deep interpretation we win by drawing a BHG (Big Heavy Green) ball. We lose by drawing a BHR (Big Heavy Red) ball. All other balls just restart the game. The graph for the deep-deep interpretation is given in Fig. 8.

The equation is also simple:

$$P_{\text{START}} = P(\text{BHG}) + rP_{\text{START}}$$

where $r = P(S) + P(\text{BL}) + P(\text{BHW}) = 1 - (P(\text{BHG}) + P(\text{BHR}))$.⁶²

So finally:

$$P_{\text{deep}}^*(B \rightarrow [H \rightarrow (\neg W \rightarrow G)]) = \frac{P(\text{BHG})}{P(\text{BHG}) + P(\text{BHR})}$$

As expected, it is equal to the conditional probability $P(\text{G}|\text{BHW}^c)$ in the initial probability space $S = (\Omega, \Sigma, P)$. Both the Import-Export Principle and the $P^*\text{CCP}$ principle hold under the deep-deep interpretation. The formula generalizes to conditionals of arbitrary length, like $A_1 \rightarrow (A_2 \rightarrow (\dots \rightarrow (A_n \rightarrow B) \dots))$. The graph for the deep-deep-...-deep interpretation is given in Fig. 9.

The equation for the graph is:

⁶¹ So now we have 12 elementary events in the initial space $S = (\Omega, \Sigma, P)$, depending on the size (Big/Small), mass (Heavy/Light) and color (White, Green, Red) of the balls:

BHW—Big, Heavy, White

BHG—Big, Heavy, Green

...

SLR—Small, Light, Red

⁶² S—the small ball, i.e. formally $S = \text{SHW} \cup \text{SHG} \cup \text{SHR} \cup \text{SLW} \cup \text{SLG} \cup \text{SLR}$; BL—the Big Light ball, i.e. $\text{BL} = \text{BLW} \cup \text{BLG} \cup \text{BLR}$. So, r is the probability of drawing any ball different from BHG, BHR.

Fig. 8 The graph for $B \rightarrow [H \rightarrow (\neg W \rightarrow G)]$ under the deep-deep interpretation

Fig. 9 The graph for $A_1 \rightarrow (A_2 \rightarrow (\dots \rightarrow (A_n \rightarrow B) \dots))$

$P_{\text{START}} = P(A_1 \dots A_n B) + P((A_1 \dots A_n)^c) P_{\text{START}}$, which means that:

$$P_{\text{START}} = \frac{P(A_1 \dots A_n B)}{P(A_1 \dots A_n)}$$

So finally:

$$P_{\text{deep}}^*(A_1 \rightarrow (A_2 \rightarrow (\dots \rightarrow (A_n \rightarrow B) \dots))) = \frac{P(A_1 \dots A_n B)}{P(A_1 \dots A_n)} = P(B | A_1 A_2 \dots A_n)$$

Obviously, the Import-Export Principle also works:⁶³

$$P_{\text{deep}}^*(A_1 \rightarrow (A_2 \rightarrow (\dots \rightarrow (A_n \rightarrow B) \dots))) = P^*((A_1 \wedge A_2 \wedge \dots \wedge A_n) \rightarrow B)$$

7.2 The shallow-shallow interpretation

The same procedure can be applied to the shallow-shallow interpretation. Consider again the conditional

⁶³ Formally, P_{deep}^* is the probability from the space for the deep - ... - deep interpretation of the conditional $A_1 \rightarrow (A_2 \rightarrow (\dots \rightarrow (A_n \rightarrow B) \dots))$, and P^* is the probability from the space for the simple conditional $(A_1 \wedge A_2 \wedge \dots \wedge A_n) \rightarrow B$. However, we can immerse simple conditionals $A \rightarrow B$ into the space for conditionals of arbitrary length by conditionalizing by a tautology T , so that $A \rightarrow B$ becomes $T \rightarrow (T \rightarrow (\dots (T \rightarrow (A \rightarrow B) \dots)))$, and in this case we will have $P^*(A \rightarrow B) = P_{\text{deep}}^*(T \rightarrow (T \rightarrow (\dots (T \rightarrow (A \rightarrow B) \dots)))) = P(B|A)$.

Fig. 10 The graph for $B \rightarrow [H \rightarrow (\neg W \rightarrow G)]$ under the shallow-shallow interpretation. The thick line (leading from MASS-COLORS to COLORS) represents the transfer from the “first-internal” game $H \rightarrow (\neg W \rightarrow G)$ to the “internal-internal” game $\neg W \rightarrow G$

If the ball is Big, then if it is Heavy, then if it is not-White, it is Green

i.e. $B \rightarrow [H \rightarrow (\neg W \rightarrow G)]$

but this time we assume that both conditionals are shallow, which means that

- i. Drawing S restarts the game.
- ii. Drawing BL redirects us to the $H \rightarrow (\neg W \rightarrow G)$ game (interpreted in the shallow way).
- iii. Drawing BHW redirects us directly to the $\neg W \rightarrow G$ game.
- iv. Drawing BHG leads to victory.
- v. Drawing BHR leads to loss.

The graph is more complex (Fig. 10).

The respective probability can be computed by writing down the system of linear equations with three variables: P_{START} , $P_{\text{MASS-COLORS}}$, P_{COLORS} . The corresponding probability space $S^* = (\Omega^*, \Sigma^*, P^*)$ consists (intuitively speaking) of all scenarios

leading from START to any of the absorbing states WIN, LOSS. It can be constructed using the general definition given in Sect. 5.4.⁶⁴

7.3 Mixed interpretations

In the last paragraph, we considered the deep-deep and shallow-shallow interpretations (also for conditionals of arbitrary length). We can produce appropriate graphs for conditionals of arbitrary length. In general, for conditionals containing n sentences, there are 2^{n-2} possible interpretations, and consequently 2^{n-2} corresponding graphs. For long conditionals this is of no philosophical importance as conditionals of length, for instance, 200 are not natural. But examples of conditionals of length 4 are natural, both in everyday communication, and, for instance, in juristic texts, where hypothetical situations, variants and scenarios that depend on many conditions are commonplace. So in this section we focus on conditionals of length four, i.e. of the form $A_1 \rightarrow (A_2 \rightarrow (A_3 \rightarrow B))$. The results can be easily generalized to conditionals of arbitrary length.

For the purpose of the discussion in this paragraph, take the symbol $\Rightarrow$ for the deep interpretation, and $\supset$ for the shallow one; $\rightarrow$ will be neutral. So, the four possible interpretations of the conditional can be written as:

1. $A_1 \Rightarrow (A_2 \Rightarrow (A_3 \rightarrow B))$, i.e. deep-deep;
2. $A_1 \Rightarrow (A_2 \supset (A_3 \rightarrow B))$, i.e. deep-shallow;
3. $A_1 \supset (A_2 \Rightarrow (A_3 \rightarrow B))$ i.e. shallow-deep;
4. $A_1 \supset (A_2 \supset (A_3 \rightarrow B))$ i.e. shallow-shallow.

If we use the *If the ball is Big, then if it is Heavy, then if it is not-White, it is Green* example, i.e. $B \rightarrow [H \rightarrow (\neg W \rightarrow G)]$, we can explain these four interpretations by fixing the rules of the game in different ways:

1. $B \Rightarrow [H \Rightarrow (\neg W \rightarrow G)]$ (deep-deep): **one** ball, which has to be Big, Heavy and Non-White—and we win if this particular ball is Green.
2. $B \Rightarrow [H \supset (\neg W \rightarrow G)]$ (deep-shallow): if we get a Big Heavy ball, then the game (provided it has not been settled) changes into a Purely Colorful Game $W \rightarrow G$.
3. $B \supset [H \Rightarrow (\neg W \rightarrow G)]$ (shallow-deep): after a Big ball has been drawn, the game (provided it has not been settled) is continued as $H \Rightarrow (\neg W \rightarrow G)$ (i.e. deeply interpreted).
4. $B \supset [H \supset (\neg W \rightarrow G)]$ (shallow-shallow): a Big ball initiates the $[H \supset (\neg W \rightarrow G)]$ game (interpreted in the shallow way).

(1) and (4) have been discussed in Sects. 8.1 and 8.2.

The corresponding graphs for these interpretations are obtained easily, and computing the probabilities is straightforward. Figure 11 contains two graphs: the

⁶⁴ $\Omega^* = \{S^n(\text{BHR}), S^n(\text{BHG}), S^n(\text{BHW})W^kR, S^n(\text{BHW})W^kG, S^n(\text{BL})L^m(\text{HR}), S^n(\text{BL})L^m(\text{HG}), S^n(\text{BL})L^m(\text{HW})W^kR, S^n(\text{BL})L^m(\text{HW})W^kG; n, m, k \geq 0\}$. The probabilities P^* of the elementary events are obtained by the standard rule, i.e. by multiplying the probabilities of the events from the sample space $S = (\Omega, \Sigma, P)$, and the extension to Σ^* is straightforward. The computations within S^* are much more tedious than by using the equations in the graph.

Fig. 11 The graphs for: $B \Rightarrow [H \supset (\neg W \rightarrow G)]$ (deep-shallow interpretation—left) and for $B \supset [H \Rightarrow \neg W \rightarrow G]$ (shallow-deep interpretation—right)

left is for $B \Rightarrow [H \supset (\neg W \rightarrow G)]$ i.e. deep-shallow; the right is for $B \supset [H \Rightarrow (\neg W \rightarrow G)]$ i.e. shallow-deep.

The probabilities can be computed by solving the appropriate systems of equations for graphs in Fig. 11.

Remember that the Stalnaker Bernoulli spaces approach imposes the shallow interpretation. McGee's account is concerned with the deep interpretation, so none of these accounts can deal with "mixed interpretations" of the conditionals. But mixed interpretations are not just a formal game, consider for instance the sentence:

If John had broken his leg in the mountains in 2017, then if we went hiking, it would hurt if he had a heavy rucksack.

The first conditional is shallow (we do **not** mean continuing his hiking with the broken leg), but the second is deep: hiking **with** a heavy rucksack would cause trouble (obviously **not** hiking on Monday and having a heavy rucksack on Tuesday). So, the structure is:

$$B \supset [H \Rightarrow (R \rightarrow L)]$$

B —John broke his leg

H —John went hiking

R —John had a heavy rucksack

L —The leg hurts.

This situation cannot, to our knowledge, be handled in the models known from the literature.

8 Formal models for conditionals revisited

In this chapter we shortly summarize our findings concerning the status of the Import-Export Principle and PCCP (9.1), discuss some methodological issues (9.2) and compare our model with the Stalnaker Bernoulli model (9.3).

8.1 IE and PCCP

How does the model solve important questions concerning conditionals?

8.1.1 The Import-Export Principle

This has already been discussed in Sect. 5. It is known that the Import-Export Principle works for the deep interpretation, i.e.

$$P_{\text{deep}}^*(A \rightarrow (B \rightarrow C)) = P_{\text{deep}}^*(B \rightarrow (A \rightarrow C)) = P^*((A \wedge B) \rightarrow C) = P(C|AB)$$

For the "multiple-deep" interpretation of the nested conditional, the Import-Export Principle also works (see Sect. 6):

$$P_{\text{deep}}^*(A_1 \rightarrow (A_2 \rightarrow (\dots \rightarrow (A_n \rightarrow B) \dots))) = P^*((A_1 \wedge A_2 \wedge \dots \wedge A_n) \rightarrow B) = P(B|(A_1 A_2 \dots A_n)).$$

For the shallow interpretation, the Import-Export Principle is not true, which also is intuitive: fulfilling the first condition changes the rules of the game, so depending on what the first condition is, the probabilities can differ greatly. So, generally:

$$P_{\text{shallow}}^*(A \rightarrow (B \rightarrow C)) \neq P_{\text{shallow}}^*(B \rightarrow (A \rightarrow C))$$

The graph model corresponds to McGee's model in the sense that it allows to account for the deep interpretation. It also allows to account for the shallow interpretation, sharing this feature with the Stalnaker Bernoulli model. However, its advantage is that it allows to account for both these interpretations at once, in particular allowing to describe mixed conditionals (Sect. 8.2).

8.1.2 PCCP

For the deep interpretation, the formula

$$P_{\text{deep}}^*(A \rightarrow (B \rightarrow C)) = P(C|AB)$$

is true, so P*CCP holds in this version. For the shallow interpretation, we have the following formula:

$$P_{\text{shallow}}^*(A \rightarrow (B \rightarrow C)) = P(BC|A) + P(B^c|A)P(C|B)$$

The components on the right side are all conditional probabilities from the initial, simple probability space. The equation retains the spirit of the simple version of P*CCP: it makes it possible to compute the probability of the nested conditional as a function (a combination) of conditional probabilities from the initial space.⁶⁵

The results concerning PCCP obtained within this article correspond to the results obtained within the other known models. Its superiority consists in the fact, that it can be generalized to longer and mixed conditionals.

8.2 Methodological aspects

Our aim is to give a philosophically justified account of conditionals and we hope that graph model meets these criteria. But in the present study we focus on the methodological advantages. In Sect. 2.2 we mentioned generalizability and simplicity in this context. These notions are not entirely clear, and important issues to address in this context arise: (i) Should the account for right-nested conditionals be an extension of the rule for simple conditionals? (ii) What are the prospects of the final theory (of probabilities of conditionals) being coherent? (iii) How is simplicity to be understood?⁶⁶

In order to make an assessment, we will indicate some features of our model and compare it with the Stalnaker Bernoulli model (which was a very important source of inspiration for us). However, regardless of the advantages of the Stalnaker

⁶⁵ So, $P_{\text{shallow}}^*(A \rightarrow (B \rightarrow C)) = \Psi(P(BC|A), P(B^c|A), P(C|B))$, where $\Psi(x, y, z) = x + yz$.

⁶⁶ We thank the anonymous referee for formulating these questions in a clear way and for the enlightening observations on these issues.

Bernoulli models and the importance and depth of van Fraassen's and Kaufmann's results, we think that modeling the conditional in terms of **finite** scenarios and graphs rather than sets of infinite paths is more intuitive.

8.2.1 Generalizability and flexibility

For simple conditionals $A \rightarrow B$, the strategy is to build a (simple) graph representing the flow of the game. When both A and B occur, the game is won; for A and $\neg B$, it is lost. For $\neg A$, the bets are not settled, and the game is restarted. So, loosely speaking, the general rule is:

(RESTART) *If the antecedent is not true—then restart!*

Dealing with conditionals of the form $A \rightarrow (B \rightarrow C)$ amounts to building a (more complex) graph as well. Just as in the simple case, if the antecedent A is false, the game is restarted. If both A and $B \rightarrow C$ occur (i.e. the conditional $B \rightarrow C$ happens to be true), the game $A \rightarrow (B \rightarrow C)$ is won. If A and $\neg(B \rightarrow C)$ occur (i.e. the conditional $B \rightarrow C$ happens to be false), the game $A \rightarrow (B \rightarrow C)$ is lost. The difference with respect to the simple case $A \rightarrow B$ is that now it might happen that A is true, whereas $B \rightarrow C$ does not have a settled status (think of Tuesday in the MED example). In this situation the bet in the game $A \rightarrow (B \rightarrow C)$ is not settled and we have to decide what to do now. The decision depends on the interpretation of the conditional. Under the **deep** interpretation, the whole game is restarted (we might say that the fact that A has taken place is forgotten and we start our observation anew). Under the **shallow** interpretation, A is considered to be decided once and forever, and we restart only the game $B \rightarrow C$. So, the general mechanism of creating the graph for $A \rightarrow (B \rightarrow C)$ is similar to the simple case $A \rightarrow B$: we also have to make a decision as to how the game is to be continued, and depending on our interpretation we fix the appropriate rules (and the graph). The RESTART rule for building graphs for $A \rightarrow (B \rightarrow C)$ is therefore a natural extension of the rule for $(A \rightarrow B)$. In this sense, the theory for right-nested conditionals might be considered to be an extension of the model for simple conditionals.⁶⁷

Observe also that the graphs for the nested conditional $A \rightarrow (B \rightarrow C)$ contain a subgraph for the simple conditional $(B \rightarrow C)$ so that we can “track” the simple (sub)conditional within the graph for the compound conditional. So, the model for more complex cases is a straightforward extension of the model for simpler cases.

This does not yet prove the claim that the final theory (still under construction) will be coherent. The final theory might use quite different methods to handle different cases. Such situations happen in mathematics, even if mathematicians usually struggle for an elegant, uniform theory. However, we think that the fact that it is possible

- i. to model both the deep and shallow interpretations of right-nested conditionals,
- ii. to model longer conditionals which allow mixed (deep/shallow) interpretations,

⁶⁷ Consider a more general conditional $A \rightarrow \beta$, with arbitrary β . The RESTART rule should hold for any β . We have shown how to apply it in a concrete way to right-nested conditionals.

- iii. to incorporate into the model both the global and local readings of the simple conditional (see Wójtowicz and Wójtowicz 2019)

provides a good motivation to work on the general theory and gives hope that it can be constructed.

8.2.2 Formal simplicity

We can graphically represent right-nested conditionals and—thanks to the properties of Markov graphs—computing the probabilities is straightforward and simple (as only solving a system of linear equations is needed). If we also think of the problem in practical terms—for instance, if we are interested in analyzing decision-making processes (based on evaluating the probabilities of conditionals)—it is important to have a method which is also of practical in addition to theoretical interest.⁶⁸ The Stalnaker Bernoulli model is more general in the sense of the possibility of formally representing diverse conditionals in one space, but with more complicated methods this comes with the price of computing the corresponding probabilities.

However, the problem of whether the Markov graph account is conceptually simpler than the Stalnaker Bernoulli account is subtle. The notion of conceptual simplicity has a pragmatic flavor and obviously depends on the criteria. Our probability space is countable, so the probability distribution on elementary events extends to arbitrary events by a simple rule. By contrast, the Stalnaker Bernoulli space has cardinality of the continuum, so the σ -field has to be constructed with care as it cannot be taken to be the power set of Ω . The graph model is ontologically simpler in the sense that finite structures are involved and the universe of constructed objects, i.e. graphs, is countable, i.e. it is smaller than the universe of the Stalnaker Bernoulli space. However, this is not a decisive criterion as in mathematics this kind of conceptual simplicity can be misleading.⁶⁹

However, regardless of the possibility of giving a general criterion of simplicity, we think that our model has the following advantages:

⁶⁸ We suppose that it might be difficult to model complex conditionals in the Stalnaker Bernoulli space because the corresponding event (i.e. the set of infinite sequences in the Stalnaker Bernoulli space) might become very complicated and the computations tedious. Also, modeling the local versus global interpretations of the simple conditional might be tedious, but in the graph model it is rather simple (for details see Wójtowicz and Wójtowicz 2019).

⁶⁹ Problems in finite combinatorics or number theory can be much more intricate than in abstract set theory. Consider the well-known simple problem: are there any natural numbers x, y, z, n (with $x, y, z, n > 0; n > 2$) such that $x^n + y^n = z^n$...? Solving Fermat's hypothesis required the use of very sophisticated algebraic geometry techniques, and the methods are very distant from the elementary notions needed to formulate the problem. By contrast, problems involving very abstract set-theoretical notions can have simple solutions. Measuring the conceptual complexity of the mathematical notions involved is a subtle problem: for instance, in Reverse Mathematics (cf. Simpson 2009) the needed resources are evaluated in terms of the strength of set existence axioms. Stronger axioms might be ontologically more costly but conceptually simpler (proofs are often easier when stronger assumptions are available). On the other hand, formulating an elementary proof (even of an already known theorem) can be an important contribution: the Fields medal in 1950 was awarded to Selberg for giving an elementary proof of the Prime Number Theorem.

- i. it provides a **simple method of giving a formal explication** of interpreting the conditional, and the possibility of giving a graphical representation is also very useful;⁷⁰
- ii. **computing the probabilities is much simpler** (solving systems of linear equations).

Colloquially speaking, we show both graphs to an agent who is in doubt about the interpretation of the conditional $A \rightarrow (B \rightarrow C)$ and ask, “which game corresponds to what you have in mind?” When the decision has been met, we show the simple method of computing the probability of the conditional in question. We think that choosing between two graphs is intuitive and easy to understand for the agent. By contrast, using the Stalnaker Bernoulli model will enforce the identification of the appropriate (infinite) class of infinite sequences. Moreover, the Stalnaker Bernoulli model enforces the shallow interpretation, therefore the agent who has the deep interpretation in mind will have first to “translate” the conditional in question (eliminating the deeply interpreted conditional) and then identify the appropriate class of sequences for the translated conditional.

8.3 Comparison with Stalnaker Bernoulli spaces

Is our model or the Stalnaker Bernoulli model more general? Consider any set of non-conditional sentences $\{A_1, \dots, A_n\}$. Every finite sequence can be imbedded into the Stalnaker Bernoulli universe in a natural way, for instance by “gluing” it with the set of all infinite sequences formed from the sentences $\{A_1, \dots, A_n\}$, i.e. the cylindrical set $\prod_{k \in \mathbb{N}} X_k$, where $X_k = \{A_1, \dots, A_n\}$. The Stalnaker Bernoulli space might therefore be considered to contain all possible sequences, i.e. all possible scenarios of all possible games. Seen from this angle, the Stalnaker Bernoulli model looks more general: every set of scenarios generated by a graph is represented in the Stalnaker Bernoulli space. In the Stalnaker Bernoulli space, all right-nested conditionals can be modeled within one space, but the construction of the space imposes the **shallow** interpretation. The deep interpretation requires a different model, and mixed conditionals are out of reach.

Importantly, in our model there is no one single space representing **all** conditionals. It is exactly the other way around: **all** conditionals are represented by respective graphs (and corresponding probability spaces). So, in some sense the Stalnaker Bernoulli model is more universal in terms of modeling all conditionals within one space. On the other hand, the graph model is more flexible with respect to identifying and explaining interpretations. It also offers more “user-friendly” computational recipes.

⁷⁰ There is an ongoing discussion in philosophy of mathematics concerning the epistemic role of diagrams (for instance: Brown 1999; Giaquinto 2007, 2008). Some authors even claim that diagrams can play the role of proofs in some cases (consider the case of category theory, where often an important idea is expressed by the fact that certain diagrams are commuting). Even if such claims might be considered exaggerated, it is true that diagrams play a big heuristic role in mathematics. Viewed from this angle, the possibility of giving a graphical representation that exhibits the interpretation of the conditional is an additional advantage.

9 Summary

We have given a model in which it is possible to assign a probability P^* to every right-nested conditional of the form $(A_1 \rightarrow (A_2 \rightarrow (\dots \rightarrow (A_n \rightarrow B) \dots)))$, depending on the interpretation of the conditional (i.e. whether the connectives are given the deep or the shallow meaning). $A_1, \dots, A_n, B$ are (non-conditional) sentences which are interpreted as events in the initial probability space $S = (\Omega, \Sigma, P)$. This general result obviously covers the simplest but philosophically very important case of the conditional $A \rightarrow (B \rightarrow C)$.

This aim has been achieved by defining appropriate graphs: for the conditional $\alpha = A \rightarrow (B \rightarrow C)$ in question, we have formally defined two graphs $G_{\alpha\text{-deep}}$ and $G_{\alpha\text{-shallow}}$ (Figs. 4 and 6) that correspond to the deep and the shallow interpretations. These graphs represent the “logico-causal” structure of the respective interpretation.

The graphs $G_{\alpha\text{-deep}}$ and $G_{\alpha\text{-shallow}}$ are associated with two different probability spaces S_{deep}^* , S_{shallow}^* . In both these spaces, the conditional $A \rightarrow (B \rightarrow C)$ has its interpretation as an event. Of course, these events are different because they correspond to different (deep versus shallow) readings of the conditional (and moreover they live in different probability spaces). The probabilities P_{deep}^* and P_{shallow}^* of the conditional $A \rightarrow (B \rightarrow C)$ computed in the respective spaces S_{deep}^* , S_{shallow}^* (Sects. 6.1, 6.2) coincide with Kaufmann’s results for shallow and with McGee’s results for the deep interpretation.

A very important feature of the model is that the probability of α can be computed in a straightforward way by directly solving the appropriate system of linear equations for the graph G_α , which is usually much simpler than direct computations within the appropriate probability space. The simplification is substantial if we consider more complex conditionals.

So, taking into account the analyses above, we claim that—at least when right-nested conditionals are considered—the presented model is more versatile and universal than the models known from the literature.

Acknowledgements The preparation of this paper was supported by the National Science Centre (NCN) Grant 2016/21/B/HS1/01955. For helpful discussions and critical comments, we would like to thank the audiences in Munich, Madrid and Dubrovnik. We would like to thank the anonymous referees whose detailed and insightful comments helped us to improve presentation of the paper. Of course, we are responsible for any mistakes.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adams, E. W. (1965). On the logic of conditionals. *Inquiry*, 8, 166–197.
- Adams, E. W. (1970). Subjunctive and indicative conditionals. *Foundations of Language*, 6, 89–94.
- Adams, E. W. (1975). *The logic of conditionals*. Dordrecht: Reidel.
- Adams, E. W. (1998). *A primer of probability logic*. Stanford, CA: CSLI Publications.
- Bertsekas, D. P. & Tsitsiklis, J. N. (2008). *Introduction to probability*, 2nd edition. Nashua, NH: Athena Scientific.
- Bod R., Hay J., Jannedy S. (2003) *Probabilistic linguistics*. Cambridge, MA: MIT Press.
- Bradley, R. (2012). Multidimensional possible-world semantics for conditionals. *Philosophical Review*, 122(4), 539–571.
- Brown, J. R. (1999). *Philosophy of mathematics: An introduction to the world of proofs and pictures*. New York: Routledge.
- Douven, I. (2016). On de Finetti on iterated conditionals. In C. Beierle, G. Brewka, & M. Thimm (Eds.), *Computational models of rationality* (pp. 265–279). Rickmansworth: College Publications.
- Douven, I., & Verbrugge, S. (2013). The probabilities of conditionals revisited. *Cognitive Science*, 37, 711–730.
- Edgington, D. (1991). Do conditionals have truth-conditions? In Jackson, F. (Ed.), *Conditionals* (pp. 176–201). Oxford: Oxford University Press.
- Edgington, D. (1995). On conditionals. *Mind*, 104, 235–329.
- Égre, P., Rossi, L., & Sprenger, J. (2019). De Finettian logics of indicative conditionals. arxiv.org/pdf/1901.10266.pdf (forthcoming in *Journal of Philosophical Logic*).
- Feller, W. (1968). *An introduction to probability theory and its applications*, vol. 1, 3rd edition. New York: Wiley.
- Giaquinto, M. (2007). *Visual thinking in mathematics*. Oxford: Oxford University Press.
- Giaquinto, M. (2008). Visualizing in mathematics. In P. Mancosu (Ed.), *The philosophy of mathematical practice* (pp. 22–42). Oxford: Oxford University Press.
- Gibbard, A. (1981). Two recent theories of conditionals. In W. Harper, R. C. Stalnaker, G. Pearce (Eds.), *Ifs: Conditionals, belief, decision, chance and time* (pp. 211–247). Dordrecht: Reidel.
- Grinstead, C. M., & Snell, J. L. (1997). *Introduction to probability*. Providence, RI: AMS.
- Hajek, A. (2011). Triviality pursuit. *Topoi*, 30(1), 3–15.
- Hajek, A. (2012). The Fall of “Adams’ Thesis”? *Journal of Logic, Language and Information*, 21, 145–161.
- Kaufmann, S. (2004). Conditioning against the grain: Abduction and indicative conditionals. *Journal of Philosophical Logic*, 33(6), 583–606.
- Kaufmann, S. (2005). Conditional predictions: A probabilistic account. *Linguistics and Philosophy*, 28(2), 181–231.
- Kaufmann, S. (2009). Conditionals right and left: Probabilities for the whole family. *Journal of Philosophical Logic*, 38, 1–53.
- Kaufmann, S. (2015). Conditionals, conditional probability, and conditionalization. In H. Zeevat, & H. C. Schmitz (Eds.), *Bayesian natural language semantics and pragmatics* (pp. 71–94). Berlin: Springer.
- Khoo, J. (2015). On indicative and subjunctive conditionals. *Philosopher’s Imprint*, 15(32), 1–40.
- Khoo, J. (2016). Probabilities of conditionals in context. *Linguistics and Philosophy*, 39, 1–43.
- Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85, 297–315.
- McGee, V. (1989). Conditional probabilities and compounds of conditionals. *Philosophical Review*, XCVIII, 4, 485–541.
- Ramsey, F. P. (1929). General propositions and causality. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays* (pp. 237–255). London: Keagan Paul, Trench, Trubner, & Co., 1931. Reprinted in: *Philosophical Papers*, Mellor, D. H. (Ed.) (pp. 145–163), Cambridge: Cambridge University Press, 1990.
- Simpson, S. (2009). *Subsystems of second order arithmetic*. Cambridge: Cambridge University Press.
- Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (pp. 98–112). Oxford: Basil Blackwell.
- Stalnaker, R. (1970). Probability and conditionals. *Philosophy of Science*, 37(1), 64–80.

- Stalnaker, R. & Jeffrey, R. (1994). Conditionals as random variables. In E. Eells, B. Skyrms (Eds.), *Probabilities and conditionals: Belief revision and rational decision* (pp. 31–46). Cambridge: Cambridge University Press.
- Stalnaker, R., & Thomason, R. H. (1970). A semantic analysis of conditional logic. *Theoria*, 36(1), 23–42.
- Van Fraassen, B. C. (1976). Probabilities of conditionals. In W. L. Harper, R. Stalnaker, & G. Pearce (Eds.), *Foundations of probability theory, statistical inference, and statistical theories of science* (pp. 261–308). Dordrecht: Reidel.
- Wójtowicz, A., & Wójtowicz, K. (2019). A stochastic graphs semantics for conditionals. *Erkenntnis*, 1–35. <https://doi.org/10.1007/s10670-019-00144-z>.
- Wójtowicz, A., & Wójtowicz, K. (in preparation). Mathematical versus practical probabilities of conditionals: a stochastic graph approach. Ms., University of Warsaw.
- Wójtowicz, A., Wójtowicz, K. (2021). Dutch Book against Lewis. *Synthese*. <https://doi.org/10.1007/s11229-021-03199-0>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.