Learning Conditional Information by Jeffrey Imaging on Stalnaker Conditionals

Mario Günther

October 30, 2017

Abstract

We propose a method of learning indicative conditional information. An agent learns conditional information by Jeffrey imaging on the minimally informative proposition expressed by a Stalnaker conditional. We show that the predictions of the proposed method align with the intuitions in Douven (2012)'s benchmark examples. Jeffrey imaging on Stalnaker conditionals can also capture the learning of uncertain conditional information, which we illustrate by generating predictions for the Judy Benjamin Problem.

Keywords. Learning Conditional Information, Stalnaker Conditional, Imaging, Douven's Examples, Judy Benjamin Problem.

Contents

1 Introduction

"How do we learn conditional information?" Igor Douven et al. present this question for consideration in a series of papers.[1] Douven (2012) contains a survey of the available accounts that model the learning of conditional information. The survey comes to the conclusion that a general account of probabilistic belief updating by learning (uncertain) conditional information is still to be formulated. Douven and Pfeifer (2014) analyses the state of the art even more pessimistically by writing that "no one seems to have an idea of what an even moderately general rule of updating on conditionals might look like", even if we restrict the scope of the account to indicative conditionals.[2]

Douven (2012) dismisses the Stalnaker conditional as a means to model the learning of conditional information. He argues for the dismissal by pointing out that the Stalnaker conditional "makes no predictions at all about any of our examples".[3] Douven provides three possible worlds models for his point. Each model consists of four worlds such that all logical possibilities of two binary variables are covered. He observes that imaging on "If α, then γ" interpreted as a Stalnaker conditional has different effects: in model I the probability of the antecedent α, i. e. $P(\alpha)$ decreases, in model II $P(\alpha)$ remains unchanged, and in model III $P(\alpha)$ increases. According to Douven this flexibility of the class of possible worlds models is a problem rather than an advantage, since there were no rationality constraints to rule out certain models as rational representations of a belief state.

Pace Douven, we show that his dismissal of the Stalnaker conditional is unjustified by proposing an updating method based on the Stalnaker semantics and inspired by Lewis's imaging method. The core idea of the proposed method is that an agent learns conditional information by Jeffrey imaging on the minimally informative meaning of the corresponding Stalnaker conditional. The method models the three examples Douven takes as benchmark for an account of learning conditional information. In addition, Jeffrey imaging, our generalisation of Lewis's imaging method, accounts for the learning of uncertain conditional information, as we will illustrate by applying our learning method to Bas Van Fraassen (1981)'s Judy Benjamin Problem.

In Section 2, we propose our probabilistic method of learning indicative conditional

[1] Cf. Douven and Dietz (2011), Douven and Romeijn (2011), Douven (2012), Douven and Pfeifer (2014, especially Section 6).

[2] Douven and Pfeifer (2014, p. 213).

[3] Douven (2012, p. 247).

information. First, we introduce the concepts of a Stalnaker conditional, Lewis's imaging, and our generalisation thereof. Based on these concepts, we supplement the properties of a Stalnaker model's similarity order by the minimally informative interpretation of a Stalnaker conditional and a default assumption. We justify both by the rationale that belief changes should be as conservative as possible. We show that the supplemented Stalnaker models provide sufficient constraints to model the learning of indicative conditional information by applying the learning method to Douven's examples as well as the Judy Benjamin Problem. Thereby we recover possible worlds approaches from Douven's dismissal.

2 A Probabilistic Method of Learning Indicative Conditional Information

The proposed learning method may be summarised as follows. (i) We model an agent's belief state as a Stalnaker model. (ii) The agent learns conditional information by (ii).(a) interpreting the received conditional information as a Stalnaker conditional, (ii).(b) constraining the similarity order by the meaning of the Stalnaker conditional in a minimally informative way and in presence of the default assumption, and (ii).(c) updating her degrees of belief by Jeffrey imaging on this Stalnaker conditional (together with further contextual information, if available). (iii) We check whether or not the result of (Jeffrey) imaging complies with the correct intuitions associated with the scenario under consideration.

In Section 2.1, we introduce the meaning of a Stalnaker conditional. In Section 2.2, we present Lewis's updating method called 'imaging', which relates the probability of a Stalnaker conditional and the probability of the consequent after imaging on the antecedent. In Section 2.3, we generalise Lewis's imaging in order to model cases of learning uncertain conditional information. In Section 2.4, we describe our method of learning conditional information in more detail, before we apply the method, in Section 2.6, to Douven's examples and the Judy Benjamin Problem. In Section 2.5, we provide a rationale for the minimally informative interpretation of Stalnaker conditionals and the default assumption for learning conditional information.

2.1 The Stalnaker Conditional

The idea behind a Stalnaker conditional may be expressed as follows: a Stalnaker conditional $\alpha > \gamma$ is true at a world w iff γ is true in the most similar possible world w' to w, in which α is true.[4] The evaluation of a Stalnaker conditional requires a model of possible worlds. A model of possible worlds, in turn, requires the specification of a logical language.

Definition 1. Full Conditional Language
Let $Prop$ be the set of atomic propositions. Then $\mathcal{L}$ be a set of formulas such that

(i) for each $p_1, p_2, ... \in Prop$, $p_i \in \mathcal{L}$,

(ii) if $\alpha, \gamma \in \mathcal{L}$, then $\neg\alpha \in \mathcal{L}$ and $\alpha \wedge \gamma \in \mathcal{L}$,

(iii) if $\alpha, \gamma \in \mathcal{L}$, then $\alpha > \gamma \in \mathcal{L}$,

(iv) and no other expressions are in $\mathcal{L}$.

We say that $\mathcal{L}$ is the full conditional language.

The full conditional language $\mathcal{L}$ contains any type of Boolean combination of conditionals, e.g. $(\alpha > \gamma) \wedge \beta \in \mathcal{L}$, and arbitrary nestings of conditionals, e.g. $\alpha > (\gamma \wedge (\beta > \delta))$.

Let w denote a Boolean assignment, or equivalently a possible world. We denote the set of Boolean assignments, or equivalently the set of possible worlds, that satisfy a formula α by $[\alpha]$. We thus identify the set $[\alpha]$ with the proposition expressed by α. In symbols, $[\alpha] = \{w \in W \mid w(\alpha) = 1\}$.

Definition 2. Stalnaker Model
We say that $\mathcal{M}_{St} = \langle W, R, \leq, V\rangle$ is a Stalnaker model iff

(i) W is a non-empty set of possible worlds,

(ii) $R : W \times W$ is a binary accessability relation over worlds such that:

  (a) for all $w \in W$: wRw. (Reflexivity)

[4] Cf. Stalnaker (1975). Note that Stalnaker's theory of conditionals aims to account for both indicative and counterfactual conditionals. We set the complicated issue of this distinction aside in this paper. However, we want to emphasise that Douven's examples and the Judy Benjamin Problem only involve indicative conditionals.

(iii) $\leq$ assigns each $w \in W$ a total order $\leq_w$ such that:

  (a) for all $w, w', w'' \in W$: if $w' \leq_w w''$ and wRw'', then wRw'.
  (b) for all $w, w' \in W$: $w' \leq_w w$, only if $w' = w$. (Unique Center Assumption)
  (c) for all $w, w', w'' \in W$: $w' \leq_w w''$ or $w'' \leq_w w'$. (Connectivity)
  (d) for all $w, w' \in W$: if wRw' for some $w' \in [\alpha] \subseteq W$, then there is a $w'' \in [\alpha]$ such that $w'' <_w w'''$ for all $w''' \in [\alpha]$.[5] We say that w'' is the unique α-world minimal under $\leq_w$. In symbols, $w'' = min_{\leq_w}[\alpha]$. (Stalnaker's Uniqueness Assumption)

(iv) V is an evaluation function of the full conditional language $\mathcal{L}$ iff

  (a) $\forall p \in Prop, \forall w \in W : V(p, w) = 1$ iff $w(p) = 1$ iff $w \models p$,
  (b) and $\forall \alpha, \gamma$ and $\forall w \in W$:
    i. $w \models \neg\alpha$ iff $w \not\models \alpha$
    ii. $w \models \alpha \wedge \gamma$ iff $w \models \alpha$ and $w \models \gamma$
    iii. $w \models \alpha > \gamma$ iff $min_{\leq_w}[\alpha] \models \gamma$ if there is a $min_{\leq_w}[\alpha]$.

We comment on two aspects of Definition 2. (I) We presented an equivalent variant to Stalnaker's original models that emphasizes the similarity order $\leq$ between possible worlds, instead of a world selection function.[6] We interpret the world $w' = min_{\leq_w}[\alpha]$ as the most similar α-world from w. (II) The accessability relation R is connective due to (iii).(a), (iii).(c), and (iii).(d) of Definition 2.

Now, we can state more precisely the meaning of a Stalnaker conditional. "If α, then γ" denotes according to Stalnaker's proposal the set of worlds (or equivalently the proposition) containing each world whose most similar α-world is a world that satisfies γ. In symbols, $[\alpha > \gamma] = \{w \mid w \models \alpha > \gamma\} = \{w \mid min_{\leq_w}[\alpha] = \emptyset$ or $min_{\leq_w}[\alpha] \models \gamma\}$.

Finally, note that any Stalnaker model validates the principle called 'Conditional Excluded Middle' according to which $(\alpha > \gamma) \vee (\alpha > \neg\gamma)$. The reason is that, for any $w \in W$, the single most similar α-world $min_{\leq_w}[\alpha]$ is either a γ-world, or else a $\neg\gamma$-world. This principle will come in handy when modeling the learning of conditional information with uncertainty. In Section 2.6.4, we will apply our method to the learning of uncertain conditional information. First, however, we introduce Lewis's imaging method and our generalisation thereof.

[5]Here as elsewhere in the paper, the strict relation $w' <_w w''$ is defined as $w' \leq_w w''$ and $w'' \not\leq_w w'$.
[6]For Stalnaker's presentation of his semantics see Stalnaker and Thomason (1970).

2.2 Lewis's Imaging

We present David Lewis's probabilistic updating method called 'imaging'.[7] We introduce a notational shortcut: for each world w and each (possible) antecedens α, $w_\alpha = min_{\leq_w}[\alpha]$ be the most similar world of w such that $w_\alpha(\alpha) = 1$. Invoking the shortcut, we can then specify the truth conditions for Stalnaker's conditional operator $>$ as follows.

$$w(\alpha > \gamma) = w_\alpha(\gamma), \text{ if } \alpha \text{ is possible.}^8 \tag{1}$$

Definition 3. Probability Space over Possible Worlds
We call $\langle W, \wp(W), P\rangle$ a probability space over a finite set of possible worlds W iff

(i) $\wp(W)$ is the set of all subsets of W,

(ii) and $P : \wp(W) \mapsto [0, 1]$ is a probability measure, i.e.

   (a) $P(W) = 1, P(\emptyset) = 0$,
   (b) and for all $X, Y \subseteq W$ such that $X \cap Y = \emptyset$, $P(X \cup Y) = P(X) + P(Y)$.

As before, we conceive of the elements of $\wp(W)$ as propositions. We define, for each α, $P(\alpha) = P([\alpha])$. We see that W corresponds to an arbitrary tautology denoted by $\top$ and $\emptyset$ to an arbitrary contradiction denoted by $\bot$. Definition 3 allows us to understand a probability measure P as a probability distribution over worlds such that each w is assigned a probability $P(w) > 0$, and $\sum_w P(w) = 1$. We may determine the probability of a formula α by summing up the probabilities of the worlds at which the formula is true.[9]

$$P(\alpha) = \sum_w P(w) \cdot w(\alpha) \tag{2}$$

Now, we are in a position to define Lewis's updating method of imaging.

[7] Cf. Lewis (1976).

[8] We assume here that there are only finitely many worlds. Note also that if α is possible, then there exists some w_α.

[9] We assume here that each world is distinguishable from any other world, i. e. for two arbitrary worlds, there is always a formula in $\mathcal{L}$ such that the formula is true in one of the worlds, but false in the other. In other words, we consider no copies of worlds.

Definition 4. Imaging (Lewis (1976, p. 310))
For each probability function P, and each possible formula α, there is a probability function P^α such that, for each world w', we have:

$$P^\alpha(w') = \sum_w P(w) \cdot \left\{ \begin{array}{ll} 1 & \text{if } w_\alpha = w' \\ 0 & \text{otherwise} \end{array} \right\} \tag{3}$$

We say that we obtain P^α by imaging P on α, and call P^α the image of P on α.

Intuitively, imaging transfers the probability of each world w to the most similar α-world w_α. Importantly, the probabilities are transfered, but in total no probability mass is additionally produced and no probability mass is lost. In formal terms, we have always $\sum_{w'} P^\alpha(w') = 1$. Any α-world w' keeps at least its original probability mass (since then $w_\alpha = w'$), and is possibly transfered additional probability shares of $\neg\alpha$-worlds w iff $min_{\leq_w}[\alpha] = w'$. In other words, each α-world w' receives as its updated probability mass its previous probability mass plus the previous probability shares that were assigned to $\neg\alpha$-worlds w such that $min_{\leq_w}[\alpha] = w'$. In this way, the method of imaging distributes the whole probability onto the α-worlds such that $P^\alpha(\alpha) = \sum_{w(\alpha)=1} P(w(\alpha)) = 1$, and each share remains 'as close as possible' at the world at which it has previously been located. For an illustration see Figure 1.

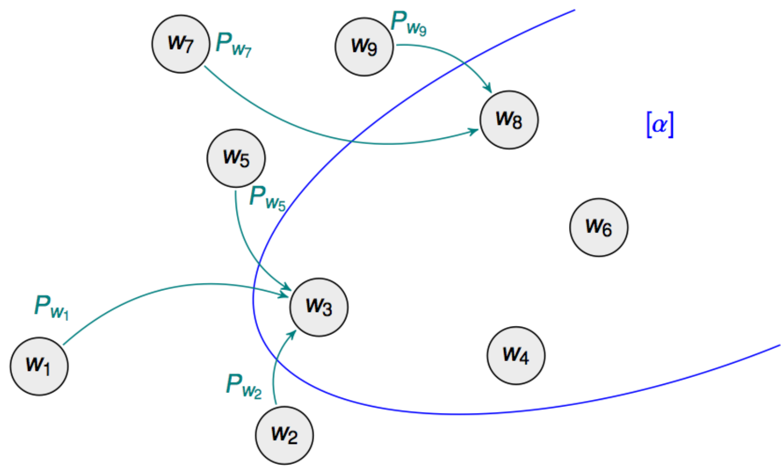

Figure 1: A set of possible worlds. The blue area represents the proposition or set of worlds $[\alpha] = \{w_3, w_4, w_6, w_8\}$. The teal arrows represent the transfer of probability shares from the respective $[\neg\alpha]$-worlds to their most similar $[\alpha]$-world. Similarity is graphically represented by topological distance between the worlds such that w_3, for instance, is the most similar or 'closest' $[\alpha]$-world to w_2.

Lewis proved the following theorem, which will be useful to model the respective examples.

Theorem 1. (Lewis (1976, p. 311))
The probability of a Stalnaker conditional equals the probability of the consequent after imaging on the antecedent, i. e. $P(\alpha > \gamma) = P^{\alpha}(\gamma)$, if α is possible.

Note that α in Theorem 1 may itself be of conditional form $\beta > \delta$ for any $\beta, \delta \in \mathcal{L}$.

2.3 Jeffrey Imaging

For the case of learning uncertain conditional information, we need to generalise Lewis's imaging method of Definition 4. In analogy to Jeffrey conditionalisation, we call the generalised method 'Jeffrey' imaging.[10] Jeffrey imaging is based on

[10] Cf. Jeffrey (1965). In personal communication, Benjamin Eva and Stephan Hartmann mentioned that the idea behind Jeffrey imaging is already used in artificial intelligence research to model the

Lewis's imaging and the fact that in a Stalnaker model the principle of Conditional Excluded Middle prescribes that $\neg(\alpha > \gamma) \equiv \alpha > \neg\gamma$. We know, for all $w \in W$, presupposed $\alpha > \gamma$ is possible, both (I) that $\sum_w P^{\alpha>\gamma}(w)$ sums up to 1 and (II) that $\sum_w P^{\alpha>\neg\gamma}(w)$ sums up to 1. The idea is that if we form a weighted sum over the terms of (I) and (II) with some parameter $k \in [0, 1]$, then we obtain again a sum of terms $P_k^{\alpha>\gamma}(w)$ such that $\sum_w P_k^{\alpha>\gamma}(w) = 1$. Note, however, that we present the more general case $P_k^{\alpha}(w)$ in the definition below.

Definition 5. Jeffrey Imaging
For each probability function P, each possible formula $\alpha \in \mathcal{L}$ (possibly of conditional form $\beta > \delta$), and some parameter $k \in [0, 1]$, there is a probability function P_k^{α} such that, for each world w' and the two similarity orderings centred on w_α and $w_{\neg\alpha}$, we have:

$$P_k^{\alpha}(w') = \sum_w (P(w) \cdot \left\{ \begin{array}{ll} k & \text{if } w_\alpha = w' \\ 0 & \text{otherwise} \end{array} \right\} + P(w) \cdot \left\{ \begin{array}{ll} (1-k) & \text{if } w_{\neg\alpha} = w' \\ 0 & \text{otherwise} \end{array} \right\}) \quad (4)$$

We say that we obtain P_k^{α} by Jeffrey imaging P on $\alpha \in \mathcal{L}$, and call P_k^{α} the Jeffrey image of P on α. Note that in the case where $k = 1$, Jeffrey imaging reduces to Lewis's imaging.

Theorem 2. Properties of Jeffrey Imaging

(i) $\sum_{w'} P_k^{\alpha}(w') = 1$

(ii) $P_k^{\alpha}(\alpha) = k$

(iii) $P_k^{\alpha}(\neg\alpha) = (1 - k)$

(iv) $P_k^{\alpha}(\gamma) = k \cdot P(\alpha > \gamma)$

retrieval of information. Sebastiani (1998, p. 3) mentions the name 'Jeffrey imaging' without writing down a corresponding formula. Crestani (1998, p. 262) says that Sebastiani (1998) suggested "a new variant of standard imaging called *retrieval by Jeffrey's logical imaging*". However, the formalisation of Jeffrey's idea on p. 263 differs from mine in at least two respects. (i) An additional truth evaluation function occurs in the formalisation for determining whether a formula (i. e. 'query') is true at a world (i. e. 'term'). (ii) Instead of a parameter k locally gouverning the probability kinematics of each possible world, Crestani simply uses a global constraint on the posterior probability distribution.

Proof. (i)

$$\begin{aligned}\sum_{w'} P_k^\alpha(w') = \sum_{w'}\sum_{w}(P(w)\cdot \left\{\begin{array}{ll} k & \text{if } w_\alpha = w' \\ 0 & \text{otherwise}\end{array}\right\} + \\ P(w)\cdot \left\{\begin{array}{ll} (1-k) & \text{if } w_{\neg\alpha} = w' \\ 0 & \text{otherwise}\end{array}\right\}) \qquad & \text{Algebra} \\ = \sum_{w'}(k\cdot \sum_{w} P(w)\cdot \left\{\begin{array}{ll} 1 & \text{if } w_\alpha = w' \\ 0 & \text{otherwise}\end{array}\right\} + \\ (k-1)\cdot \sum_{w} P(w)\cdot \left\{\begin{array}{ll} 1 & \text{if } w_{\neg\alpha} = w' \\ 0 & \text{otherwise}\end{array}\right\}) \qquad & \text{by Def. 4, } \sum_{w'}\sum_{w} P(w) = 1 \\ = k + (k-1) = 1 \qquad & \text{for all } k \in [0,1]\end{aligned} \tag{5}$$

(ii)

$$\begin{aligned}P_k^\alpha(\alpha) = \sum_{w_\alpha} P_k^\alpha(w_\alpha)\cdot w_\alpha(\alpha) \qquad & \text{Definition 5} \\ = \sum_{w_\alpha}(\sum_{w} P(w)\cdot \left\{\begin{array}{ll} k & \text{if } w_\alpha = w_\alpha \\ 0 & \text{otherwise}\end{array}\right\} + \\ \sum_{w} P(w)\cdot \left\{\begin{array}{ll} (1-k) & \text{if } w_{\neg\alpha} = w_\alpha \\ 0 & \text{otherwise}\end{array}\right\})\cdot w_\alpha(\alpha) \qquad & \text{Second term cancels out} \\ = \sum_{w} P(w)\sum_{w_\alpha}\cdot \left\{\begin{array}{ll} k & \text{if } w_\alpha = w_\alpha \\ 0 & \text{otherwise}\end{array}\right\}\cdot w_\alpha(\alpha) \qquad & \text{Algebra} \\ = k\cdot \sum_{w} P(w)\cdot w_\alpha(\alpha) \qquad & \text{Algebra and (2)} \\ = k\cdot P^\alpha(\alpha) = k\cdot P(\alpha > \alpha) = k\end{aligned} \tag{6}$$

(iii) Obvious.

(iv) Replace in (ii) the ‘consequent’ α by γ.

□

We see that in total the revision method of Jeffrey imaging does neither produce additional probability shares, nor destroy any probability shares. In contrast to Lewis’s imaging, Jeffrey imaging does not distribute the whole probabilistic mass onto the α-worlds, but only a part thereof that is determined by the parameter k.

In particular, as compared to Lewis's imaging, Jeffrey imaging may be understood as implementing a more moderate or balanced movement of probabilistic mass between α- and $\neg\alpha$-worlds. For an illustration see Figure 2.

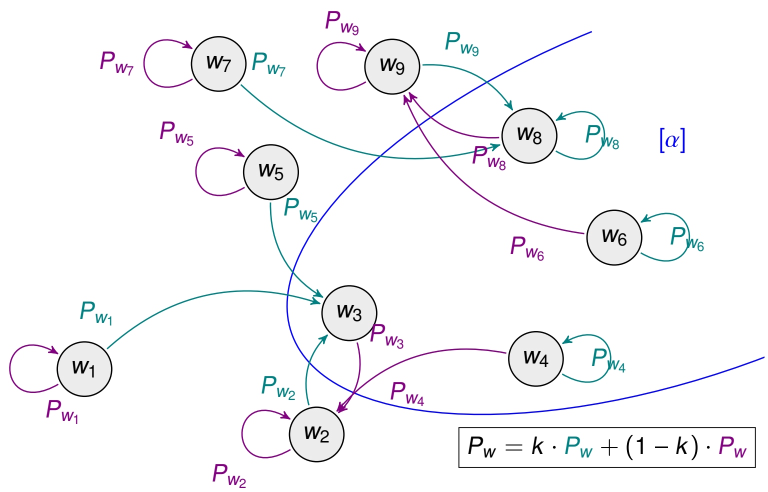

Figure 2: An illustration of the probability kinematics of Jeffrey imaging. The Jeffrey image P_k^α is characterised by a 'k-inertia' of the probabilistic mass from the respective α-worlds, and a '$(1-k)$-inertia' of the probabilistic mass from the respective $\neg\alpha$-worlds. Each teal arrow represents the transfer of the probability share $k \cdot P(w)$ to the closest α-world from w. Each violet arrow represents the transfer of the probability share $(1-k) \cdot P(w)$ to the closest $\neg\alpha$-world from w.

It is easy to show that P_k^α is a probability function. In a possible worlds framework, such a proof basically amounts to showing that the probability shares of all the worlds sum up to 1 after Jeffrey imaging. Therefore, property (i) of Theorem 2 provides minimal justification for applying Jeffrey imaging to probabilistic belief updating.

2.4 A Simple Method of Learning Conditional Information

We outline now a method of learning conditional information in three main steps.

(i) We model an agent's belief state as a Stalnaker model $\mathcal{M}_{St} = \langle W, R, \leq, V \rangle$ such that all and only those logical possibilities are represented as single worlds, which are relevant to the scenario under consideration. For instance, if only a single conditional "If α, then γ" is relevant and nothing else, then W contains exactly four elements as depicted in Figure 3.[11]

(ii) An agent learns conditional information "If α, then γ" iff (a) the agent interprets the received conditional information as a Stalnaker conditional $\alpha > \gamma$, (b) changes the similarity order $\leq$ by the meaning of $\alpha > \gamma$ in a minimally informative way and respecting the default assumption, and (c) updates her degrees of belief by Jeffrey imaging on the minimally informative meaning of $\alpha > \gamma$.

(iii) Finally, we check whether or not the result of Jeffrey imaging obtained in step (ii).(c) corresponds to the intuition associated with the respective example.

Step (ii) constitutes the core of the learning method and requires further clarification.

(a) In the agent's belief state, i. e. a Stalnaker model, the received information is interpreted. In the case of conditional information, the received information is interpreted as Stalnaker conditional. Hence, if the agent receives the information "If α, then γ", she interprets the information as meaning that the most similar α-world (from the respective actual world) is a world that satisfies γ (presupposed α is possible). Technically, the interpretation (i. e. the meaning) of $\alpha > \gamma$ relative to the Stalnaker model $\mathcal{M}_{St}$ is the proposition $[\alpha > \gamma] = \{w \in W \mid min_{\leq_w}[\alpha] \in [\gamma]\}$, where w is the respective actual world.

(b) The similarity order(s) is/are changed upon receiving conditional information. The proposition $\{w \in W \mid min_{\leq_w}[\alpha] \in [\gamma]\}$ depends on the similarity order $\leq$. The learning method prescribes that $\leq$ is specified, or adjusted, such that from each world the most similar α-world is a γ-world whenever *possible*. In other words, the method demands a maximally conservative, or equivalently minimally informative, interpretation of the received information. This amounts to specifying or adjusting the orders $\leq_w$ such that *as many worlds as possible* satisfy the received information. On the one hand, we can describe this interpretation as maximally conservative in the sense that no worlds are gratuitously excluded. On the other hand, we may think of

[11] In other words, we consider "small" possible worlds models and do not allow for copies of worlds, i. e. worlds that satisfy the same formulas.

possible worlds as information states. Then the exclusion of possible worlds corresponds to a gain of information. If an agent interprets the received information in a maximally conservative way, then as few as possible worlds or information states are excluded. In this sense, her gain of information is minimal. We will also use the abbreviation $[\alpha > \gamma]_{min}$ for the minimally informative proposition expressed by $\alpha > \gamma$.

The learning method prescribes that the agent changes her similarity order respecting a default assumption. This default assumption states that the most similar $\alpha > \gamma$-world from any excluded $\alpha > \neg\gamma$-world is a $\alpha \wedge \gamma$-world, if there is more than one candidate. Formally, this constraint expresses that $min_{\leq_{w(\alpha > \neg\gamma)=1}}[\alpha > \gamma] \vDash \alpha \wedge \gamma$, if $min_{\leq_{w(\alpha > \neg\gamma)=1}}[\alpha > \gamma]$ is underdetermined. ($min_{\leq_{w(\alpha > \neg\gamma)=1}}[\alpha > \gamma]$ denotes the respectively most similar $\alpha > \gamma$-world from any world w such that $w(\alpha > \neg\gamma) = 1$ given the minimally informative similarity order $\leq$.) Notice the interplay between the two constraints for the similarity order: the minimally informative interpretation of the Stalnaker conditional minimises the number of worlds which might have several candidates for the most similar $\alpha > \gamma$-world and, based on this interpretation, the default assumption determines the most similar $[\alpha > \gamma]$-world for each world having more than one candidate world (at least in absence of further information).

(c) Jeffrey imaging is applied on the minimally informative meaning of the Stalnaker conditional $\alpha > \gamma$. The application of Jeffrey imaging determines a probability distribution after learning the (uncertain) conditional information.

We note that the proposed learning method has the following properties resembling *modus ponens* and *modus tollens*. If the agent already knows $[\alpha]$ (or $[\neg\gamma]$ resp.), learning the minimally informative proposition $[\alpha > \gamma]$ implies that the agent also knows $[\gamma]$ (or $[\neg\alpha]$ resp.).

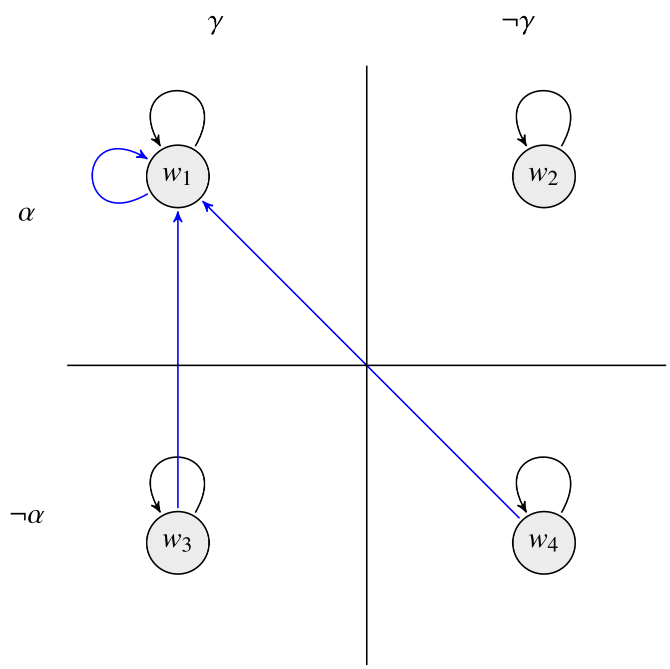

Figure 3: A four-worlds Stalnaker model for a case, in which the only received and relevant information is "If α, then γ". The reflexive arrows illustrate that each world w is the most similar to itself under the respective similarity order $\leq_w$. The blue arrows illustrate the change of the similarity order such that the received and interpreted information $[\alpha > \gamma]$ is minimally informative. Here, the minimally informative meaning of $\alpha > \gamma$ is $[\alpha > \gamma] = \{w \in W \mid w \vDash \alpha > \gamma\} = \{w_1, w_3, w_4\}$. Note that according to (iii).(b) and (iii).(d) of Definition 2 world w_2 is its own most similar α-world, but does not satisfy γ, i. e. $min_{\leq_{w_2}}[\alpha] \nvDash \gamma$ and thus $min_{\leq_{w_2}}[\alpha > \gamma] \neq w_2$. Relying on the default assumption of step (ii).(b), $min_{\leq_{w_2}}[\alpha > \gamma] = w(\alpha \wedge \gamma) = w_1$. In words, the method prescribes that w_1 is the most similar $\alpha > \gamma$-world to w_2, which is excluded under the minimally informative meaning of $\alpha > \gamma$. This illustrates that the minimally informative meaning $[\alpha > \gamma]$ implies that $\neg\gamma$ is excluded under the supposition of α. Hence, imaging on the minimally informative meaning of $\alpha > \gamma$ 'probabilistically excludes' w_2 and the probability share of w_2 will be fully transferred to w_1.

2.5 A Rationale for the Minimally Informative Interpretation and the Default Assumption

We aim to justify the minimally informative interpretation of conditional information and the default assumption by the following rationale: an agent should change her belief state as conservatively as possible when learning a proposition. We say that a belief change is as conservative as possible, or equivalently maximally conservative, iff the change is not stronger than necessary to believe the received information.[12]

We argue first that the rationale of maximally conservative belief change warrants the minimally informative interpretation of the received conditional information. The learning of an indicative conditional is constraint by its meaning. By Stalnaker's semantics, an indicative conditional $\alpha > \gamma$ means that γ is the case on the supposition of α. Hence, upon learning the indicative conditional (and having or receiving no more information), the agent does not necessarily learn whether or not α is the case, and *a fortiori* whether or not γ is the case. However, the agent learns at least that the conjunction $\alpha \wedge \neg\gamma$ is not possible. Correspondingly, the minimally informative meaning of $[\alpha > \gamma]$, or equivalently $[\alpha > \gamma]_{min} = [(\alpha \wedge \gamma) \vee (\neg\alpha \wedge \gamma) \vee (\neg\alpha \wedge \neg\gamma)]$, is exhausted by all those possible worlds that verify the material implication.[13] Consequently, $[\neg\alpha] \subset [\alpha > \gamma]_{min}$ and $[\gamma] \subset [\alpha > \gamma]_{min}$, which means that the minimally informative proposition $[\alpha > \gamma]$ is less informative than $[\neg\alpha]$ and $[\gamma]$, respectively. When learning $[\alpha > \gamma]_{min}$ and having or receiving no more information, an agent learns only that the $\alpha \wedge \neg\gamma$-world is excluded, whereas the learning remains silent on the status of the $\neg\alpha$-worlds. Learning only $[\alpha > \gamma]_{min}$ thus qualifies as a maximally conservative belief change: the agent believes the conditional information $[\alpha > \gamma]$ without believing any of the more informative propositions $[\alpha]$, $[\neg\alpha]$, $[\gamma]$ and/or $[\neg\gamma]$.

There comes a problem of underdetermination when learning the minimally informative proposition $[\alpha > \gamma]$. In general, the most similar $\alpha > \gamma$-world from any excluded $\neg(\alpha > \gamma)$-world is underdetermined, and thus it is not determined whereto the probability shares of the excluded worlds are transferred. In the Stalnaker model depicted in Figure 3, for instance, there are three candidate worlds

[12] For proposals and justifications of a similar rationale, see Gärdenfors (1988) and van Benthem and Smets (2015). For a critical and elucidating discussion of the principle of minimal or conservative belief change, see Rott (2000).

[13] Here the question may arise why we do not simply learn conditional information by Jeffrey imaging on the material implication. A short answer will be provided in the Conclusion.

for the most similar $\alpha > \gamma$-world from the excluded w_2. We resort to the default assumption (introduced in Section 2.4) to solve the problem of underdetermination.

The default assumption states that the most similar $\alpha > \gamma$-world from any excluded $\alpha > \neg\gamma$-world is a $\alpha \wedge \gamma$-world, if there is more than one candidate. Observe that all worlds included in the minimally informative proposition $[\alpha > \gamma]$ are themselves, respectively, their unique most similar $[\alpha > \gamma]$-world due to the reflexivity of the acquired similarity order. In this way the minimally informative interpretation minimises the number of excluded worlds, for which the default assumption may be needed to overcome the problem of underdetermination. In case only α and γ are relevant and all an agent learns is $[\alpha > \gamma]_{min}$, the default assumption simply states that the $\alpha \wedge \gamma$-world is the most similar world to the excluded $\alpha \wedge \neg\gamma$-world. If the agent obtains additional information over and above $[\alpha > \gamma]_{min}$, the default assumption leads to different outcomes – depending on which additional information is learned. The outcome's non-rigidity or variability with respect to different contextual information is illustrated by the Ski Trip Example and Driving Test Example studied in sections 2.6.2 and 2.6.3, respectively.

There is a link between the default assumption and the probability distribution after learning: in the case where the only received and relevant information is $[\alpha > \gamma]_{min}$, the default assumption is satisfied iff the probability of the antecedent remains unchanged. Assume the default assumption is in place and an agent does not come to believe a more informative proposition than $[\alpha > \gamma]_{min}$. Then our learning method prescribes that the probability share of the $\alpha \wedge \neg\gamma$-world is transferred to the $\alpha \wedge \gamma$-world. Hence, $P^{\alpha > \gamma}(\alpha) = P(\alpha)$. For the converse, assume $P^{\alpha > \gamma}(\alpha) = P(\alpha)$ and that all an agent learns is $[\alpha > \gamma]_{min}$. The only transferred probability share is again $P(\alpha \wedge \neg\gamma)$. Suppose for reductio that $P(\alpha \wedge \neg\gamma)$ is transferred to some $\neg\alpha$-world. Since $P(\alpha \wedge \neg\gamma) > 0$, $P^{\alpha > \gamma}(\neg\alpha)$ would be greater than $P(\neg\alpha)$. By the probability calculus, it would follow that $P^{\alpha > \gamma}(\alpha) < P(\alpha)$, which contradicts the assumption. Hence, $P(\alpha \wedge \neg\gamma)$ is transferred to the $\alpha \wedge \gamma$-world, and thus the default assumption is satisfied.

We argue now that the default assumption is warranted by the rationale of maximally conservative belief change. A consequence of this rationale is that there should be no belief change without any reason. (If there were such a belief change, the change would be stronger than necessary, and thus violating the rationale.) Upon learning the indicative conditional "If α, then γ", there seems to be no reason to change the probability of the antecedent, at least in the absence of additional information.[14] As we have seen in the previous paragraph, if the agent does not possess or receive additional information, the default assumption ensures that the probability of the antecedent remains unchanged when learning "If α, then γ" interpreted as $[\alpha > \gamma]_{min}$. In the absence of further information, the default assumption thus implements a demand of maximal conservativeness, viz. that the probability of the antecedent should remain unchanged.

Let us consider cases where further contextual information is given. Here, the default assumption does not necessarily ensure that the probability of the antecedent remains unchanged. If the contextual information is sufficient to uniquely determine the respective most similar world, we do not need to rely on the default assumption. We will see below that additional (contextual) information may sometimes fully determine the epistemic states under consideration such that we are not in need of the default assumption. If we need to rely on the default assumption, in contrast, we should judge the assumption by its predictions for specific scenarios. Unfortunately, there is a myriad of possible scenarios that differ in their contextual information. However, we may refer again to the case studies in sections 2.6.2 and 2.6.3, in which our learning method generates the intuitively correct results. Hence, the default assumption seems to be justified in absence and presence of further information, at least *prima facie*. This being said, we encourage the search for counterexamples to our learning method, especially counterexamples involving contextual information.

In line with our learning method, Douven and Romeijn (2011) suggest that the probability of the antecedent does not change upon learning an indicative conditional, at least in the absence of further relevant information.[15] They write:

> We are inclined to think that Adams conditioning, or, equivalently, Jeffrey conditioning with the explicit constraint of keeping the antecedent's probability fixed in the update [...] covers most of the cases of learning a conditional. (p. 654)

Since Adams conditioning always keeps the probability of the antecedent fixed,

[14]Notice that the assumption of no additional information literally excludes that there is an epistemic reason, i. e. some belief apart from $[\alpha > \gamma]_{min}$, to change the probability of the antecedent.

[15]Douven (2012) argues more precisely that the probability of the antecedent should only change if the antecedent is explanatorily relevant for the consequent. It is noteworthy that if the probability of the antecedent should intuitively change in one of Douven's examples, the explanatory relations always involve beliefs in additional propositions (apart from the conditional) given by the example's context description.

it is no general method for learning conditional information, as sometimes this probability should change. It "may entirely fall upon us", so Douven and Romeijn (2011, p. 660), "to decide, on the basis of contextual information, whether or not [Adams conditioning] applies to the learning of a given conditional"; and further "deciding when Adams conditioning applies, may be an art, or a skill, rather than a matter of calculation or derivation from more fundamental epistemic principles."

In contrast to Douven and Romeijn (2011), our learning method proposes a principled way how learning conditional information should affect the probability of the antecedent. As a consequence of the minimally informative interpretation and the default assumption, learning indicative conditional information does not alter the probability of the antecedent in the absence of further information. If, in addition, contextual information is learned, the default that the probability of the antecedent remains unchanged may be violated. Given the respective contextual information, our learning method automatically calculates a possibly changed probability for the antecedent. No skilful decision about which method should be applied is required.

2.6 Douven's Examples and the Judy Benjamin Problem

We apply now our method of learning conditional information to Douven's examples and the Judy Benjamin Problem.

2.6.1 A Possible Worlds Model for the Sundowners Example

Example 1. The Sundowners Example (Douven and Romeijn (2011, pp. 645–646))
Sarah and her sister Marian have arranged to go for sundowners at the Westcliff hotel tomorrow. Sarah feels there is some chance that it will rain, but thinks they can always enjoy the view from inside. To make sure, Marian consults the staff at the Westcliff hotel and finds out that in the event of rain, the inside area will be occupied by a wedding party. So she tells Sarah:

If it rains tomorrow, we cannot have sundowners at the Westcliff. (7)

Upon learning this conditional, Sarah sets her probability for sundowners and rain to 0, but she does not adapt her probability for rain.

We model Sarah's belief state as the Stalnaker model $\mathcal{M}_{S\,t} = \langle W, R, \leq, V \rangle$ depicted in Figure 4. W contains four elements covering the possible events of $R, \neg R, S, \neg S$, where R stands for "it rains tomorrow" and S for "Sarah and Marian can have sundowners at the Westcliff tomorrow".

Sarah interprets the conditional uttered by her sister Marian as saying that the most similar R-world from the respective 'actual' world is a world that satisfies $\neg S$. Note the symmetry to the scheme depicted in Figure 3. Critics may find reasons why the default assumption in (ii).(b) of Section 2.4 is unjustified, and thus that we encounter here the problem of underdetermination. However, as Douven himself points out, the intuition in the Sundowners Example derives from the verdict that whether or not it rains may affect whether or not they have sundowners, but not the other way around: having sundowners simply has no effect whatsoever on whether or not it rains.[16] Hence, the change of belief between R and $\neg R$ is more far fetched than between S and $\neg S$. In other words, the worlds along the horizontal axis are more similar than the worlds along the vertical axis. Consequently, $min_{\leq_{w_1}}[R > \neg S] = w_2$.

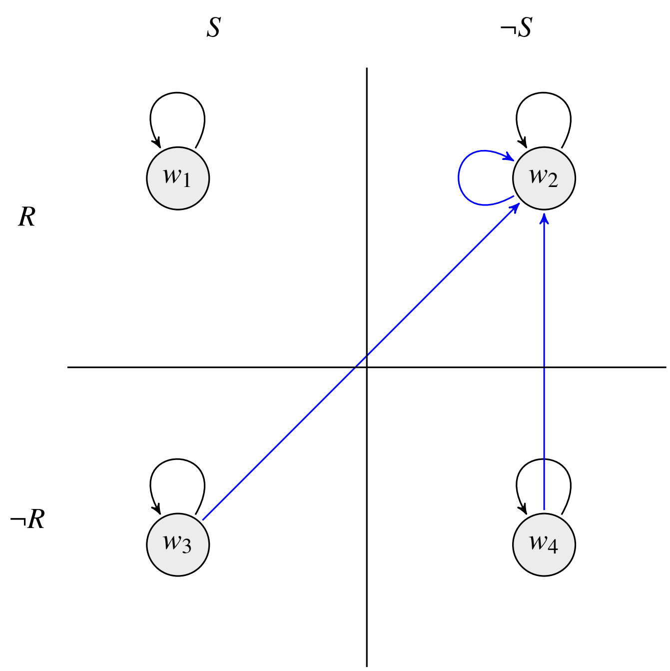

Figure 4: A Stalnaker model for Sarah's belief state in the Sundowners Example.

[16]Cf. Douven (2012, p. 8).

Imaging on the minimally informative proposition $[R > \neg S] = \{w_2, w_3, w_4\}$ results in
$P^{R>\neg S}(w') = \sum_w P(w) \cdot \left\{ \begin{array}{ll} 1 & \text{if } w_{R>\neg S} = w' \\ 0 & \text{otherwise} \end{array} \right\}$:

$$\begin{array}{ll} P^{R>\neg S}(w_1) = 0 & P^{R>\neg S}(w_2) = P(w_1) + P(w_2) \\ P^{R>\neg S}(w_3) = P(w_3) & P^{R>\neg S}(w_4) = P(w_4) \end{array} \tag{8}$$

We see immediately that both intuitions associated with the Sundowners Example are satisfied, viz. $P^{R>\neg S}(R) = P(R) = P(w_1) + P(w_2)$ and $P^{R>\neg S}(R \wedge S) = P^{R>\neg S}(w_1) = 0$. We conclude that our method yields the intuitively correct results.

Although justified by Douven's remarks, the Sundowners Example demands to invoke the default assumption in order to avoid the problem of underdetermination. We will see in the following examples that more contextual information, in particular additional factual information, may render the default assumption superfluous.[17]

2.6.2 A Possible Worlds Model for the Ski Trip Example

Example 2. The Ski Trip Example (Douven and Dietz (2011, p. 33))
Harry sees his friend Sue buying a skiing outfit. This surprises him a bit, because he did not know of any plans of hers to go on a skiing trip. He knows that she recently had an important exam and thinks it unlikely that she passed. Then he meets Tom, his best friend and also a friend of Sue's, who is just on his way to Sue to hear whether she passed the exam, and who tells him:

If Sue passed the exam, her father will take her on a skiing vacation. (9)

Recalling his earlier observation, Harry now comes to find it more likely that Sue passed the exam.

We model Harry's belief state as the Stalnaker model $\mathcal{M}_{St} = \langle W, R, \leq, V \rangle$ depicted in Figure 5. W contains eight elements covering the possible events of $E, \neg E, S, \neg S, B, \neg B$, where E stands for "Sue passed the exam", S for "Sue's father takes her on a skiing vacation", and B for "Sue buys a skiing outfit".

[17] Note that the Sundowners Example seems to be somewhat artificial. It seems plausible that upon hearing her sister's conditional, Sarah would promptly ask "why?" in order to obtain some more contextual information, before setting her probability for sundowners and rain to 0. After all, she "thinks that they can always enjoy the view from inside".

Harry interprets the conditional uttered by his friend Tom as saying that the most similar E-world from the actual world is a world that satisfies S. Crucially, Harry observed Sue buying a skiing outfit, and thus has the factual information that B.

In total, Harry learns the minimally informative proposition $[(E > S) \wedge B] = \{w \in W \mid min_{\leq_w}[E] \in [S] \wedge B\} = \{w_1, w_3, w_4\}$. Moreover, the default assumption states that $w_1(E \wedge S \wedge B) = 1$ is the most similar world to all worlds not included in the minimally informative proposition $[(E > S) \wedge B]$ (see the caption of Figure 5).

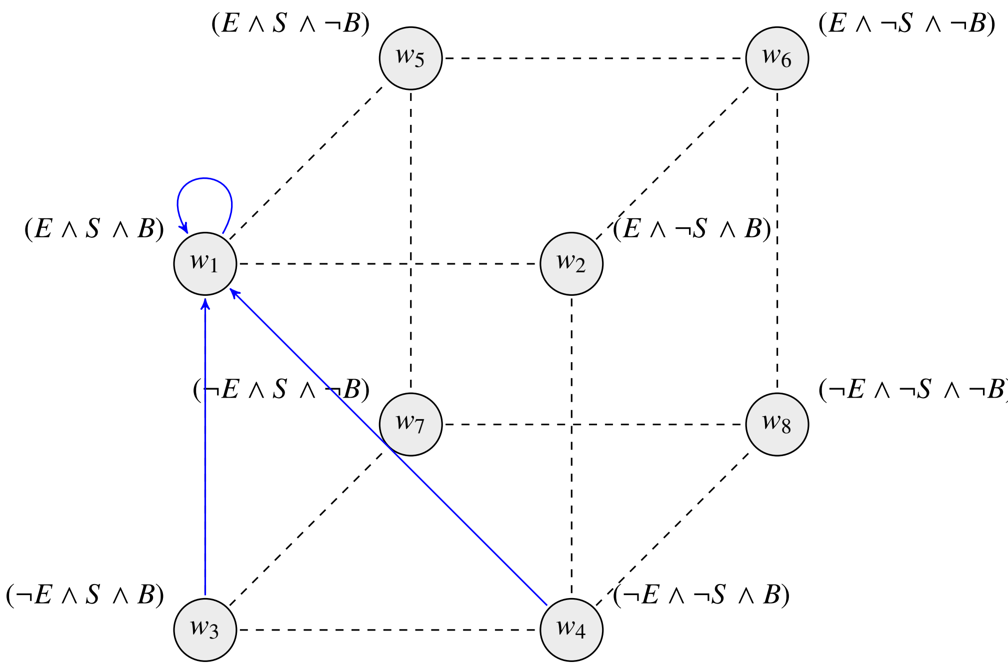

Figure 5: An eight-worlds Stalnaker model for Harry's belief state in the Ski Trip Example. The blue arrows illustrate the change of the similarity order such that the received and interpreted information $[(E > S) \wedge B)]$ is minimally informative. By the default assumption, $w_1(E \wedge S \wedge B) = 1$ is the most similar world from any $\neg((E > S) \wedge B)$-world in case of underdetermination. Note that receiving the factual information that B excludes all the worlds on the back of the cube. We see that receiving factual information is very informative, as compared to obtaining conditional information.

Imaging on the minimally informative proposition $[(E > S) \wedge B] = \{w_1, w_3, w_4\}$

results in
$P^{(E>S)\wedge B}(w') = P^*(w') = \sum_w P(w) \cdot \left\{ \begin{array}{ll} 1 & \text{if } w_{(E>S)\wedge B} = w' \\ 0 & \text{otherwise} \end{array} \right\}$:

$$\begin{array}{ll} P^*(w_1) = P(w_1) + P(w_2) + P(w_5) + P(w_6) + P(w_7) + P(w_8) & P^*(w_2) = 0 \\ P^*(w_3) = P(w_3) & P^*(w_4) = P(w_4) \\ P^*(w_5) = 0 & P^*(w_6) = 0 \\ P^*(w_7) = 0 & P^*(w_8) = 0 \end{array} \tag{10}$$

Our method yields again the correct result regarding the intuition associated with the Ski Trip Example: $P^*(E) > P(E)$, since $P^*(E) = P^*(w_1)$ and $P(E) = P(w_1) + P(w_2) + P(w_5) + P(w_6)$.

The just presented model of the Ski Trip Example is surprisingly simple. Critics may say it is too simple: your model should not omit the relation between B and S. Indeed, the first two sentences of the Ski Trip Example suggest that there is some relation between Sue's buying a skiing outfit and her going on a skiing trip (or vacation). Intuitively, (a) going on a skiing trip is a good explanation for buying a skiing outfit ($S > B$), and (b) buying a skiing outfit is a good predictor of going on a skiing trip ($B > S$). Correspondingly, the critics may say that Harry rather learns $[(E > S) \wedge (S > B) \wedge B]$ or $[(E > S) \wedge (B > S) \wedge B]$.

We will show now that our learning method's result for the Ski Trip Example is robust with respect to (a) and (b), i. e. it is compatible with assuming that Harry assumes, knows, or learns (a) or (b) in addition to $[(E > S) \wedge B]$. (Keep in mind, however, that we have no need to explicitly assume or model any relation between B and S in order to obtain the intuitively correct result.)

Let's assume Harry knows (a) $S > B$. In total, he comes to know the minimally informative proposition $[(E > S) \wedge (S > B) \wedge B] = \{w \in W \mid min_{\leq_w}[E] \in [S] \wedge min_{\leq_w}[S] \in [B] \wedge B\} = \{w_1, w_3, w_4\}$. Since the minimally informative proposition $[(E > S) \wedge (S > B) \wedge B]$ is identical to the minimally informative proposition $[(E > S) \wedge B]$, our method yields the same result. Hartmann and Rad (2017) make and need an assumption similar to (a). "From the story it is clear", they write, that it "is more likely that Sue buys a new ski outfit if her father invites her for a ski trip than if he does not" (p. 11). They represent the relation between B and S as a directed arrow from S to B in a Bayesian network claiming this would "properly represent the causal relation between these variables" (ibid.).

In contrast to the 'causal relation' between the variables, Harry could engage in the predictive inference from Sue buying a skiing outfit to raising the likelihood that she will go on a skiing trip. So, let's assume Harry thinks (b) $B > S$. In total, he comes to know the minimally informative proposition $[(E > S) \wedge (B > S) \wedge B] = \{w_1, w_3\}$. Since the minimally informative proposition $[(E > S) \wedge (B > S) \wedge B]$ is a subset of the minimally informative proposition $[(E > S) \wedge B]$, our method yields an even stronger result: if B is a predictor of S and B is believed, then it is (possibly) even more likely that Sue passed the exam (E). We obtain $P^{(E>S)\wedge(B>S)\wedge B}(E) \geq P^{(E>S)\wedge B}(E) > P(E)$. (Notice that our method is also apt to handle cases when Harry assumes, knows, or learns the conditionals (a) or (b) to a certain degree, as we will illustrate in the discussion of the Judy Benjamin Problem.)

Of course, Harry could also be equipped with some other contextual knowledge. Douven (2012, p. 11) himself, for example, provides an alternative picture:

> Sue's having passed the exam would, if true, *explain* why she bought the skiing outfit; that makes her having passed the exam more credible.

Here, Douven seems to propose that Harry has a relation between B and E in mind, viz. $E > B$ (given he knows already $E > S$ and B). On this picture, Harry learns in total the minimally informative proposition $[(E > S) \wedge (E > B) \wedge B] = \{w_1, w_3, w_4\}$. We see that if S and B are related in the sense that they are both a 'consequent' of E, our method yields again the correct result.

We take this as further evidence that our method's result for the Ski Trip Example is largely independent of and compatible with additional plausible assumptions between its variables. If the respective additional assumptions correspond to different (admissible) interpretations, the robustness of the result could explain why it is intuitively so clear that Harry should believe it more likely that Sue passed the exam.[18] In any case, our learning method needs fewer assumptions than the other accounts to obtain the desired result for the Ski Trip Example. At the same time, the result still stands if we adopt the additional assumptions on which Douven (2012) as well as Hartmann and Rad (2017) rely.

2.6.3 A Possible Worlds Model for the Driving Test Example

Example 3. The Driving Test Example (Douven (2012, p. 3))

[18] In Günther (2017), we generalise the proposed method to the learning of causal information, which allows us to define an inference to the best explanation scheme, as Douven envisioned for the Ski Trip Example.

Betty knows that Kevin, the son of her neighbors, was to take his driving test yesterday. She has no idea whether or not Kevin is a good driver; she deems it about as likely as not that Kevin passed the test. Betty notices that her neighbors have started to spade their garden. Then her mother, who is friends with Kevin's parents, calls her and tells her the following:

If Kevin passed the driving test, his parents will throw a garden party. (11)

Betty figures that, given the spading that has just begun, it is doubtful (even if not wholly excluded) that a party can be held in the garden of Kevin's parents in the near future. As a result, Betty lowers her degree of belief for Kevin's having passed the driving test.

We model Betty's belief state as the Stalnaker model $\mathcal{M}_{St} = \langle W, R, \leq, V \rangle$ depicted in Figure 6. W contains eight elements covering the possible events of $D, \neg D, G, \neg G, S, \neg S$, where D stands for "Kevin passed the driving test", G for "Kevin's parents will throw a garden party", and S for "Kevin's parents have started to spade their garden".

Betty interprets the conditional uttered by her mother as saying that the most similar D-world from the actual world is a world that satisfies G. Furthermore, Betty infers from her contextual knowledge that if Kevin's parents are spading their garden, then they will not throw a garden party, in symbols $S > \neg G$. Therefore, Betty also obtains the information that the most similar S-world from the actual world is a world that satisfies $\neg G$. Finally, Betty knows that Kevin's parents have started to spade their garden, and thus has the factual information that S.

In total, Betty learns the minimally informative proposition $[(D > G) \wedge (S > \neg G) \wedge S] = \{w \in W \mid min_{\leq_w}[D] \in [G] \wedge min_{\leq_w}[S] \in [\neg G] \wedge S\} = \{w_4\}$. The obtained information, although interpreted in a minimally informative way, is sufficient to identify the actual world. Therefore, the Stalnaker model provides us with a unique most similar world (to any other world) under the changed similarity order (see the caption of Figure 6).

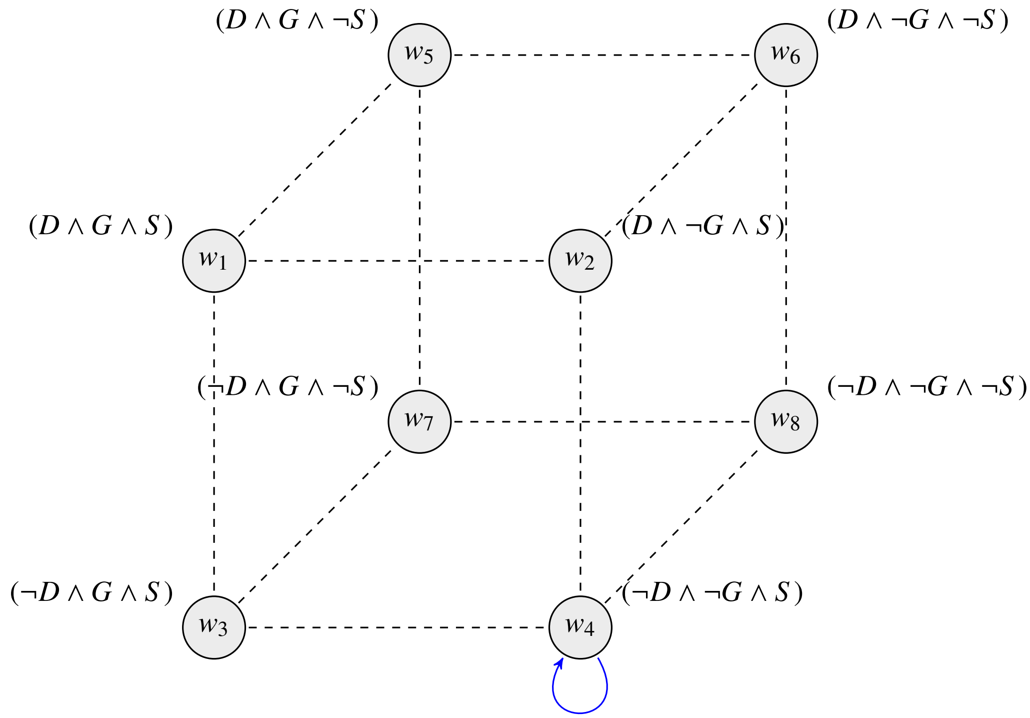

Figure 6: An eight-worlds Stalnaker model for Betty's belief state in the Driving Test Example. There is only a single world that satisfies the 'minimally informative' proposition $[(D > G) \wedge (S > \neg G) \wedge S]$. For, $[(D > G) \wedge (S > \neg G)] = \{w \in W \mid min_{\leq_w}[D] \in [G] \wedge min_{\leq_w}[S] \in [\neg G]\} = \{w_4, w_5, w_7, w_8\}$. Of those worlds only w_4 is in the S-plane of the cube and thus the actual world.

Intuitively, Betty learns that she is in a S-world, since she factually obtains the information that S. Hence, the conditional $S > \neg G$ implies that $\neg G$ is true in the actual world. By the conditional $D > G$, we know that G is satisfied in the most similar D-world from the actual world. Since $\neg G$ is true in the actual world, we know that the actual world is not a D-world. But then the actual world is a $\neg D$-world. For, if the actual world w were a D-world, w would satisfy G. To summarise, the actual world satisfies $\neg D$, $\neg G$, and, obviously, S.

Imaging on the minimally informative proposition $[(D > G) \wedge (S > \neg G) \wedge S] = \{w_4\}$ results in

$$P^{(D>G)\wedge(S>\neg G)\wedge S}(w') = P^*(w') = \sum_w P(w) \cdot \left\{ \begin{array}{ll} 1 & \text{if } w_{(D>G)\wedge(S>\neg G)\wedge S} = w' \\ 0 & \text{otherwise} \end{array} \right\}:$$

$$\begin{array}{ll} P^*(w_1) = 0 & P^*(w_2) = 0 \\ P^*(w_3) = 0 & P^*(w_4) = 1 \\ P^*(w_5) = 0 & P^*(w_6) = 0 \\ P^*(w_7) = 0 & P^*(w_8) = 0 \end{array} \tag{12}$$

Our method yields again the correct result regarding the intuition associated with the Driving Test Example: $P^*(D) < P(D)$, since $P^*(D) = 0$ and $P(D) = P(w_1) + P(w_2) + P(w_5) + P(w_6) > 0$. The following Judy Benjamin Problem will show that if Betty thinks that the conditional $D > G$ or $S > \neg G$ (or both) is/are uncertain, then the probability shares for some other worlds will not reduce to zero. This fact fits nicely with the Driving Test Examples's remark that "given the spading that has just begun, it is doubtful [or uncertain] (even if not wholly excluded) that a party can be held in the garden of Kevin's parents". We will treat the learning of uncertain conditional information in the next section.

2.6.4 A Possible Worlds Model for the Judy Benjamin Problem

We apply now our method of learning conditional information to a case, in which the received conditional information is uncertain. We show thereby that the method may be generalised to cases in which the learned conditional information is uncertain, provided we use Jeffrey imaging. Following the presentation in Hartmann and Rad (2017), we consider Bas Van Fraassen's Judy Benjamin Problem.[19]

Example 4. The Judy Benjamin Problem (Hartmann and Rad (2017, p. 7))
A soldier, Judy Benjamin, is dropped with her platoon in a territory that is divided in two halves, Red territory and Blue territory, respectively, with each territory in turn being divided in equal parts, Second Company area and Headquarters Company area, thus forming four quadrants of roughly equal size. Because the platoon was dropped more or less at the center of the whole territory, Judy Benjamin deems it equally likely that they are in one quadrant as that they are in any of the others. They then receive the following radio message:

> I can't be sure where you are. If you are in Red Territory,
> then the odds are 3 : 1 that you are in Second Company area. (13)

[19] Cf. Van Fraassen (1981, pp. 376–379).

After this, the radio contact breaks down. Supposing that Judy accepts this message, how should she adjust her degrees of belief?

Douven claims that the probability of R should, intuitively, remain unchanged after learning the uncertain conditional information. Furthermore, the probability distribution after hearing the radio message, i. e. P^*, should take the following values:

$$\begin{aligned} P^*(R \land S) = \frac{3}{8} \quad & P^*(R \land \neg S) = \frac{1}{8} \\ P^*(\neg R \land S) = \frac{1}{4} \quad & P^*(\neg R \land \neg S) = \frac{1}{4} \end{aligned} \tag{14}$$

We model Judy Benjamin's belief state as the Stalnaker model $\mathcal{M}_{St} = \langle W, R, \leq, \leq', V\rangle$ depicted in Figure 7. W contains four elements covering the possible events of $R, \neg R, S, \neg S$, where R stands for "Judy Benjamin's platoon is in Red territory", and S for "Judy Benjamin's platoon is in Second Company area". The story prescribes that the probability distribution before learning the uncertain information is given by:

$$P(R \land S) = P(R \land \neg S) = P(\neg R \land S) = P(\neg R \land \neg S) = \frac{1}{4} \tag{15}$$

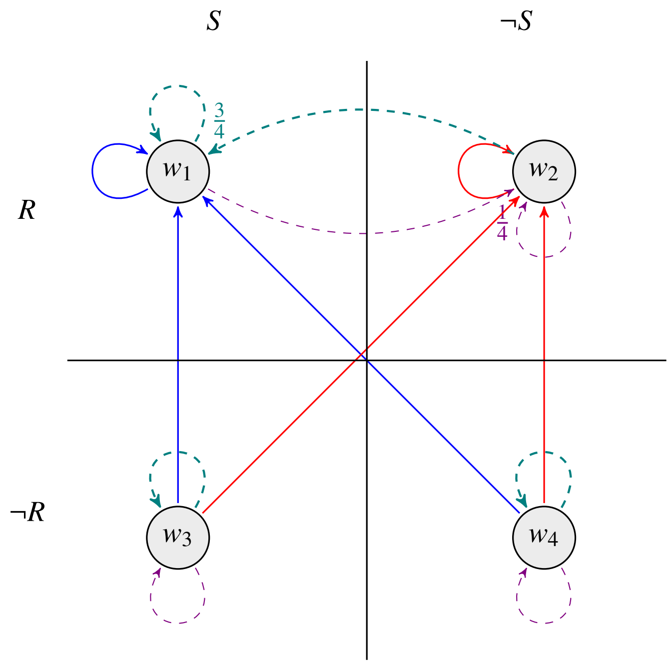

Figure 7: A Stalnaker model for Private Benjamin's belief state in the Judy Benjamin Problem. The blue arrows illustrate the specification of a similarity order $\leq'$ such that the received information $[R > S]$ is minimally informative. Note that each world having a blue arrow satisfies $R > S$. The red arrows illustrate the specification of another similarity order $\leq \neq \leq'$ such that the received information $[R > \neg S]$ is minimally informative. Each world having a red arrow satisfies $R > \neg S$. In sum, the similarity orders are specified such that one makes $[R > S]$ a minimally informative proposition and the other makes $[R > \neg S]$ a minimally informative proposition. By the default assumption, we obtain $min_{\leq_{w(R>\neg S)=1}}[R > S] = w_1$ and $min_{\leq'_{w(R>S)=1}}[R > \neg S] = w_2$. Note that under $\leq$, worlds w_3 and w_4 satisfy $R > S$, under $\leq'$ they satisfy $R > \neg S$. The teal arrows represent the transfer of $k \cdot P(w)$, while the violet arrows represent the transfer of $1 - k \cdot P(w)$. The application of Jeffrey imaging on $[R > S]$ with $k = \frac{3}{4}$ leads to the following probability distribution: $P^{R>S}_{\frac{3}{4}}(w_3) = \frac{3}{4} \cdot P(w_3) + \frac{1}{4} \cdot P(w_3)$ and $P^{R>S}_{\frac{3}{4}}(w_4) = \frac{3}{4} \cdot P(w_4) + \frac{1}{4} \cdot P(w_4)$, whereas $P^{R>S}_{\frac{3}{4}}(w_1) = \frac{3}{4} \cdot P(w_1) + \frac{3}{4} \cdot P(w_2)$ and $P^{R>S}_{\frac{3}{4}}(w_2) = \frac{1}{4} \cdot P(w_1) + \frac{1}{4} \cdot P(w_2)$.

In the previous examples, our agents learned a Stalnaker conditional $\alpha > \gamma$ with certainty. According to Theorem 1, this amounts to the constraint that $P(\alpha > \gamma) = P^{\alpha}(\gamma) = 1$ (provided α is not a contradiction). Given this constraint and since P^{α} is a probability distribution, we have $P^{\alpha}(\neg\gamma) = 1 - P^{\alpha}(\gamma) = 0$. This means that we were able to probabilistically exclude any $\neg\gamma$-world under the supposition of α.

Now, our agent Judy Benjamin learns a Stalnaker conditional with uncertainty. According to Theorem 1, this amounts in this case to the constraint that $P(R > S) = P^{R}(S) = \frac{3}{4}$. By the law of total probability, we obtain $P(R > \neg S) = P^{R}(\neg S) = \frac{1}{4}$. In contrast to learning conditional information with certainty, we cannot subtract the whole probabilistic mass from the $\neg S$-worlds under the supposition of R. However, Judy Benjamin is informed from an external source about the proportion to which she should gradually 'exclude' or downweigh the probability share of $\neg(S > R)$-worlds. Equivalently, we may say that the most similar $S > R$-world (from any $\neg(S > R)$-world) obtains a gradual upweight of probability such that it receives $\frac{3}{4}$ of the probability shares of the $\neg(R > S)$-worlds; in turn, however, this $\neg(S > R)$-world then receives a probability share from the $R > S$-world weighed by $\frac{1}{4}$. Note that in Stalnaker models $\neg(S > R) \equiv S > \neg R$ given a similarity order (and provided S is possible). Judy Bejamin's learning process may thus be modeled by considering the degree of belief in two Stalnaker conditionals that are, under a single similarity order, by the principle of Conditional Excluded Middle, mutually exclusive.

We apply now Jeffrey imaging to the Judy Benjamin Problem, where a source external to Judy provides her with the information that $k = \frac{3}{4}$.

$$P^{R>S}_{\frac{3}{4}}(w') = \sum_{w}(P(w) \cdot \begin{cases} k & \text{if } w_{R>S} = w' \\ 0 & \text{otherwise} \end{cases} + P(w) \cdot \begin{cases} (1-k) & \text{if } w_{R>\neg S} = w' \\ 0 & \text{otherwise} \end{cases}) \tag{16}$$

Given the probability distribution before the learning process in Equation (15), Judy obtains the following probability distribution after being informed that $P(R > S) = \frac{3}{4}$:

$$\begin{aligned} P^{R>S}_{\frac{3}{4}}(w_1) &= P^{R>S}_{\frac{3}{4}}(R \wedge S) = \frac{3}{8} \qquad & P^{R>S}_{\frac{3}{4}}(w_2) &= P^{R>S}_{\frac{3}{4}}(R \wedge \neg S) = \frac{1}{8} \\ P^{R>S}_{\frac{3}{4}}(w_3) &= P^{R>S}_{\frac{3}{4}}(\neg R \wedge S) = \frac{1}{4} \qquad & P^{R>S}_{\frac{3}{4}}(w_4) &= P^{R>S}_{\frac{3}{4}}(\neg R \wedge \neg S) = \frac{1}{4} \end{aligned} \tag{17}$$

Our learning method using Jeffrey imaging matches the intuitions Douven claims to be correct for the Judy Benjamin Problem. Note, in particular, that $P^{R>S}_{\frac{3}{4}}(R) = P(R) = \frac{1}{2}$.

3 Conclusion

We have seen that Douven's dismissal of the Stalnaker conditional as a tool to model the learning of conditionals is unjustified. Rather, the learning may be modelled by Jeffrey imaging on the meaning of Stalnaker conditionals under the following two conditions: (i) the similarity order of the Stalnaker model is changed in a way such that the meaning of the conditional is minimally informative, and (ii) the default assumption is in place. The proposed learning method leads to the intuitively correct results in Douven's examples and complies with Douven's intuition for the Judy Benjamin Problem.

The minimally informative meaning of a Stalnaker conditional corresponds to the meaning of the material implication. So why – for the sake of simplicity – do we not propose a method of learning conditional information by Jeffrey imaging on the material implication? There are two reasons. First, the application of Jeffrey imaging is defined with respect to similarity orders independent of using the material implication or the Stalnaker conditional. In addition, the equivalent of the default assumption for the material implication requires (an equivalent to) a similarity order as well. Hence, the proposal to Jeffrey image on the material implication is *prima facie* not more simple than our proposal.

Second, it is far from clear how to formulate Jeffrey imaging and the default assumption for the material implication with respect to uncertainty, contextual information and nested conditionals, as we did for the Stalnaker conditional. The material implication and the minimally informative interpreted Stalnaker conditional come apart when negated. Whereas the negation of the material implication carries the strong information $[\alpha \wedge \neg\gamma]$, the negation of the minimally informative proposition expressed by a Stalnaker conditional, i. e. $[\neg(\alpha > \gamma)]_{min}$, is again a minimally informative proposition, viz. $[\alpha > \neg\gamma]_{min} = [\neg\alpha \vee \neg\gamma]$. As should be clear by now, this difference is crucial for the learning of uncertain conditional information by Jeffrey imaging and shows that the Stalnaker conditional is better suited than the material implication.

Even if we try to repair the material implication account by simply stipulating that $\neg(\alpha \rightarrow \gamma)$ means $\alpha \rightarrow \neg\gamma$ (which it doesn't!), we run into further problems when considering nested conditionals. Our learning method validates the import-export principle for right-nested conditionals, i. e. $[\alpha > (\beta > \gamma)]_{min} = [(\alpha \wedge \beta) > \gamma]_{min}$, and minimally informative interpreted right-nested Stalnaker conditionals correspond to their material counterparts given their usual meaning, i. e. $[\alpha > (\beta > \gamma)]_{min} =$

$[\alpha \rightarrow (\beta \rightarrow \gamma)]$. However, for left-nested conditionals, we obtain $[(\alpha \rightarrow \beta) \rightarrow \gamma] \subset [(\alpha > \beta) > \gamma]_{min} \subset [(\alpha \wedge \beta) > \gamma)]_{min}$, a divergence between the material implication and the Stalnaker conditional that is not easily overcome.

As we have just defended against the material implication, the proposed learning method is based on Stalnaker's semantics for conditionals. We are not claiming that Stalnaker's semantics are the correct semantics for conditionals. Rather the employed semantics seem to be a good approximation. However, we are committed to the following conditional claim: if Stalnaker's semantics for conditionals are considered to be correct, or at least rational (for many cases), then the proposed learning method should be considered correct, or at least rational (for those cases).

In Günther (2017), we adapt the proposed method to the learning of uncertain causal information. The adaptation is inspired by Lewis (1973)'s notion of causal dependence for Stalnaker conditionals. This move allows us to formally implement Douven (2012)'s idea of how the degree of belief in the antecedent should change as a result of its explanatory status when learning a conditional. The combination of the methods provides a unified framework that manages to clearly discern between the more informative causal and merely conditional interpretation of a conditional.[20] Moreover, the follow-up paper includes a brief comparison between the proposed account and Douven's as well as Hartmann and Rad's account of learning conditional and/or causal information.

Acknowledgments. Thanks to Hannes Leitgeb, Stephan Hartmann, Igor Douven, and Hans Rott for helpful discussions. Special thanks go to an anonymous referee for very constructive comments. I am grateful that I had the opportunity to present parts of this paper and obtain feedback at the Munich Center for Mathematical Philosophy (LMU Munich), at the Inaugural Conference of the East European Network for Philosophy of Science (New Bulgarian University), at the International Rationality Summer Institute 2016 (Justus Liebig University), at the Center for Advanced Studies Workshop on "Learning Conditionals" (LMU Munich), at the University of Bayreuth and the University of British Columbia. This research is supported by the Graduate School of Systemic Neurosciences.

[20]This paper and Günther (2017) overlap insofar the latter contains parts of the proposed method of learning conditional information as a constituent of the adapted method. In Günther (2017), only the adapted method for learning causal information is applied to Douven's examples and the Judy Benjamin Problem; the proofs for Theorem 2 are not included.

A A Possible Worlds Model of the Jeweller Example

Following the presentation in Douven and Romeijn (2011), we consider the Jeweller Example.

Example 5. The Jeweller Example (Douven and Romeijn (2011, p. 654))
A jeweller has been shot in his store and robbed of a golden watch. However, it is not clear at this point what the relation between these two events is; perhaps someone shot the jeweller and then someone else saw an opportunity to steal the watch. Kate thinks there is some chance that Henry is the robber (R). On the other hand, she strongly doubts that he is capable of shooting someone, and thus, that he is the shooter (S). Now the inspector, after hearing the testimonies of several witnesses, tells Kate:

If Henry robbed the jeweller, then he also shot him. (18)

As a result, Kate becomes more confident that Henry is not the robber, while her probability for Henry having shot the jeweller does not change.

We model Kate's belief state as the Stalnaker model $\mathcal{M}_{St} = \langle W, R, \leq, \leq' V \rangle$ depicted in Figure 8. W contains four elements covering the possible events of $R, \neg R, S, \neg S$, where R stands for "Henry is the robber", and S for "Henry has shot the jeweller". The example suggests that $0 < P(R) < 1$ and $P(S) = \epsilon$ for a small ϵ, and thus $P(\neg S) = 1 - \epsilon$. The prescribed intuitions are that $P^*(R) < P(R)$ and $P^*(S) = P(S)$. We know about Kate's degrees of belief before receiving the conditional information that $0 < P(w_1) + P(w_2) < 1$ and $P(w_1) + P(w_3) = \epsilon$, as well as $P(w_2) + P(w_4) = 1 - \epsilon$. Note that Kate is 'almost sure' that $\neg S$, and thus we may treat $\neg S$ as 'almost factual' information.

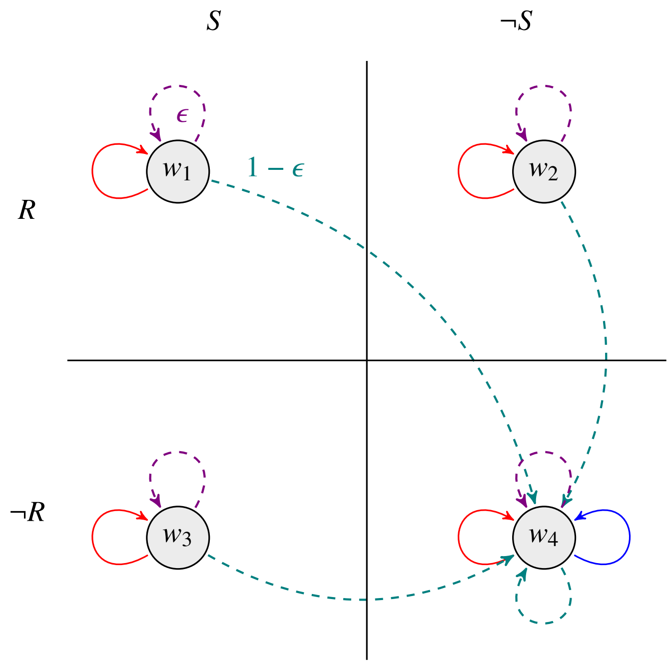

Figure 8: A Stalnaker model for Kate's belief state in the Jeweller Example. The blue arrow indicates the unique $w_{((R>S)\wedge\neg S)}$-world under $\leq$. The red arrows indicate that each world is its most similar $\neg((R > S)\wedge\neg S)$-world under $\leq'$. The teal arrows represent the transfer of $(1-\epsilon)\cdot P(w)$, while the violet arrows represent the transfer of $\epsilon \cdot P(w)$.

Kate receives certain conditional information. She learns the minimally informative proposition $[R > S] = \{w_1, w_3, w_4\}$ such that $P(R > S) = P^R(S) = 1$. By the law of total probability, $P(R > \neg S) = P^R(\neg S) = 0$. Taking her uncertain but almost factual information into account, Kate learns in total the minimally informative proposition $[(R > S) \wedge \neg S]$, which is identical to $\{w_4\}$. By $P(R > S) = 1$, $P((R > S) \wedge \neg S) = P(\neg S) = 1 - \epsilon$. Note the tension expressed in $P((R > S) \wedge \neg S) = 1 - \epsilon$. It basically says that S is almost surely not the case *and*, under the supposition of R, we exclude the possibility of $\neg S$. Intuitively, the thought expressed by this statement should cast doubt as to whether R is the case.

By $\neg((R > S)\wedge\neg S) \equiv (R > \neg S)\vee S$, we also know that $P(R > \neg S)\vee S) = \epsilon$. Note that the proposition $[(R > S) \wedge \neg S] = \{w_4\}$ (interpreted as minimally informative) specifies a similarity order $\leq$ such that $w_{(R>S)\wedge\neg S} = w_4$ for all w. In contrast, the proposition $[(R > \neg S) \vee S]$ is minimally informative in a strong sense, since it does not exclude any world w. Hence, the 'maximally inclusive' proposition $[(R > \neg S) \vee S] = \{w_1, w_2, w_3, w_4\}$ specifies a similarity order $\leq' \neq \leq$ according to which $w_{(R>\neg S)\vee S} = w$ for each w.

We apply now Jeffrey imaging to the Jeweller Example, where $k = 1 - \epsilon$.

$$P_{1-\epsilon}^{(R>S)\wedge\neg S}(w') = P^*(w') = \sum_w (P(w) \cdot \left\{ \begin{array}{ll} 1-\epsilon & \text{if } w_{(R>S)\wedge\neg S} = w' \\ 0 & \text{otherwise} \end{array} \right\} + P(w) \cdot \left\{ \begin{array}{ll} \epsilon & \text{if } w_{(R>\neg S)\vee S} = w' \\ 0 & \text{otherwise} \end{array} \right\}) \tag{19}$$

We obtain the following probability distribution after learning:

$$\begin{array}{ll} P^*_{1-\epsilon}(w_1) = P^*_{1-\epsilon}(R \wedge S) = \epsilon \cdot P(w_1) & P^*_{1-\epsilon}(w_2) = P^*_{1-\epsilon}(R \wedge \neg S) = \epsilon \cdot P(w_2) \\ P^*_{1-\epsilon}(w_3) = P^*_{1-\epsilon}(\neg R \wedge S) = \epsilon \cdot P(w_3) & P^*_{1-\epsilon}(w_4) = P^*_{1-\epsilon}(\neg R \wedge \neg S) \\ & = (1-\epsilon) \cdot (P(w_1) + P(w_2) + P(w_3) + \\ & P(w_4)) + \epsilon \cdot P(w_4) \end{array} \tag{20}$$

The results almost comply with the prescribed intuitions. The intuition concerning the degree of belief in R is met: $P^*(R) < P(R)$, since $P^*_{1-\epsilon}(w_1) + P^*_{1-\epsilon}(w_2) < P(w_1) + P(w_2)$. The intuition concerning the degree of belief in S is 'almost' met: $P^*_{1-\epsilon}(S) \approx P(S)$, for $P(w_1) + P(w_3) = \epsilon$ and $P^*_{1-\epsilon}(w_1) + P^*_{1-\epsilon}(w_3) = P(w_1) \cdot \epsilon + P(w_3) \cdot \epsilon \approx \epsilon$. In words, the method gives us the result that Kate is now pretty sure that Henry is neither the shooter nor the robber.

References

Crestani, F. Logical imaging and probabilistic information retrieval. In Crestani, F., Lalmas, M., and van Rijsbergen, C. J., editors, *Information Retrieval: Uncertainty and Logics: Advanced Models for the Representation and Retrieval of Information*, pages 247–279. Springer US, Boston, MA, 1998.

Douven, I. Learning conditional information. *Mind & Language*, 27(3):239–263, 2012.

Douven, I. and Dietz, R. A puzzle about stalnaker's hypothesis. *Topoi*, 30(1): 31–37, 2011.

Douven, I. and Pfeifer, N. Formal epistemology and the new paradigm psychology of reasoning. *Review of Philosophy and Psychology*, 5:199–221, 2014.

Douven, I. and Romeijn, J.-W. A new resolution of the judy benjamin problem. *Mind*, 120(479):637–670, 2011.

Gärdenfors, P. *Knowledge in Flux*. MIT Press, Cambridge, Mass., 1988.

Günther, M. Learning conditional and causal information by jeffrey imaging on stalnaker conditionals. *Organon F*, 24(4):456–486, 2017.

Hartmann, S. and Rad, S. R. Learning indicative conditionals. *Unpublished manuscript*, pages 1–28, 2017.

Jeffrey, R. C. *The Logic of Decision*. Mc Graw-Hill, New York, 1965.

Lewis, D. K. Causation. *Journal of Philosophy*, 70(17):556–567, 1973.

Lewis, D. K. Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85(3):297–315, 1976.

Rott, H. Two dogmas of belief revision. *Journal of Philosophy*, 97:503–522, 2000.

Sebastiani, F. Information retrieval, imaging and probabilistic logic. *Computers and Artificial Intelligence*, 17(1):1–16, 1998.

Stalnaker, R. C. A theory of conditionals. In Sosa, E., editor, *Causation and Conditionals*, pages 165–179. OUP, 1975.

Stalnaker, R. C. and Thomason, R. H. A semantic analysis of conditional logic. *Theoria*, 36(1):23–42, 1970.

van Benthem, J. and Smets, S. Dynamic logics of belief change. In Ditmarsch, H. V., Halpern, J. Y., der Hoek, W. V., and Kooi., B., editors, *Handbook of Logics for Knowledge and Belief*, chapter 7, pages 299–368. College Publications, 2015.

Van Fraassen, B. C. A problem for relative information minimizers in probability kinematics. *The British Journal for the Philosophy of Science*, 32(4):375–379, 1981.