Philosophical Logic
Counterfactuals, the standard theory

Frank Veltman

1 Counterfactual conditionals

Counterfactual conditionals are sentences of the form

$$\ulcorner \textit{If it had been the case that } \varphi, \textit{ it would have been the case that } \psi \urcorner \qquad \text{(a)}$$

They are typically uttered in contexts where the antecedent is false and known to be false. Therefore, they cannot be analyzed as material implications, because material implications with a false antecedent are true no matter what the consequent says.

Counterfactuals cannot be analyzed as strict implications either. One cannot equate a sentence of the form given in a with a formula of the form

$$\Box(\varphi \to \psi) \qquad \text{(b)}$$

where $\Box$ is the necessity operator of any normal system of modal logic, because any such system validates logical principles that do not hold for counterfactuals. One such principle is *Strengthening the Antecedent*. In any extension of **K**, we have

$$\Box(\varphi \to \chi) \models \Box((\varphi \wedge \psi) \to \chi)$$

However, from

If I had put sugar in my coffee, it would have tasted better, (c)

it does not follow that

If I had put sugar and diesel oil in my coffee, it would have tasted better. (d)

Starting point for the discussion in the following sections is the analysis of counterfactuals developed by Robert Stalnaker [20] and David Lewis [16]. Roughly put, they proposed the following truth condition for counterfactual conditionals.

- A sentence of the form ⌜*If it had been the case that φ, it would have been the case that ψ*⌝ is true in the actual world w iff the consequent ψ is true in all accessible worlds in which (a) the antecedent φ is true, and which (b) in other respects differ minimally from w.

In other words, the consequent ψ need not be true in *all* accessible worlds in which the antecedent φ is true, which it would have to be if counterfactuals were strict implications. What matters is ψ's truth value in a particular subset of this set, the φ-worlds that are most similar to the actual world.

It is easy to see how this semantics blocks the inference from (c) to (d). Consider the set S of worlds in which (i) *I put sugar in my coffee* is true and which (ii) in other respects differ minimally from the actual world. Presumably, *I put diesel oil in my coffee* is false in all these worlds. Given this, the set T of worlds in which (i) *I put sugar and diesel oil in my coffee* is true, but which (ii) in other respects differ minimally from the actual world will not be a subset of S. Now, *the coffee tastes fine* could very well be true in every world in S, but false in some of the worlds in T.

2 The System

Let us get more precise. In the sequel we are interested in languages, frames and models that are built up as follows.

- Extend the languages of propositional logic with a new binary operator $\rightsquigarrow$. Until further notice we will read '$\varphi \rightsquigarrow \psi$' as '*If it had been the case that φ, it would have been the case that ψ*'.
- Interpret the resulting languages in frames $\mathfrak{F} = \langle W, \prec \rangle$, where (i) $W \neq \emptyset$ and (ii) $\prec$ is a function which assigns to every $w \in W$ a strict partial ordering $\prec_w$ on some subset W_w of W. The elements of W will play the role of possible worlds. Until further notice the strict partial ordering $\prec_w$ is meant to play the role of a comparative similarity relation; read '$u \prec_w v$' as 'u is more similar to w than v'. The field W_w of this relation $\prec_w$ is the set of worlds that are accessible from w. Inaccessible worlds, i.e. the worlds outside W_w are supposed to be so unlike w that in w it is absurd to assume that the real world might have been be one of those.
- Supply a frame with a valuation V which assigns a truth value to every atomic sentence in every world to get a model $\mathfrak{M} = \langle W, \prec, V \rangle$. As elsewhere in this book, '$\mathfrak{M}, w \models \varphi$' is used to indicate that the formula φ is true in the world w (of the model $\mathfrak{M}$). I will write '$[\![\varphi]\!]_{\mathfrak{M}}$' to refer to $\{w \in W \mid \mathfrak{M}, w \models \varphi\}$, and call this set the proposition expressed by φ (in $\mathfrak{M}$). When it is clear which model $\mathfrak{M}$ is at stake the subscript '$\mathfrak{M}$' in $[\![\varphi]\!]_{\mathfrak{M}}$ will be omitted. Worlds in $[\![\varphi]\!]_{\mathfrak{M}}$ will be called $[\![\varphi]\!]$-worlds.
- Add the following clause to the list of truth conditions for the standard connectives.

$\mathfrak{M}, w \models \varphi \rightsquigarrow \psi$ iff for every $u \in W_w \cap [\![\varphi]\!]$ the following holds:
there is some $u' \in [\![\varphi]\!]$ such that $u' \preceq_w u$ and $\mathfrak{M}, u'' \models \psi$ for every $u'' \in [\![\varphi]\!]$ such that $u'' \preceq_w u'$.[1]

Part of the complexity of this truth condition is due to the fact that the partial orders introduced above do not have to satisfy the so-called

Limit Assumption : For every $w \in W$, the relation $\prec_w$ is well-founded.

Call any $u \in U$ a *closest* U-world to w iff $u \in W_w \cap U$ and there is no $v \in U$ such that $v \prec_w u$. Given the Limit Assumption we can be sure that in every non empty subset U of W_w we can find some worlds that are closest to w. This enables us to reformulate the truth condition in a more perspicuous way.

- Suppose the frame $\mathfrak{F} = \langle W, \prec \rangle$ satisfies the Limit Assumption, and consider the model $\mathfrak{M} = \langle W, \prec, V \rangle$. The following holds:
  $\mathfrak{M}, w \models (\varphi \rightsquigarrow \psi)$ iff $\mathfrak{M}, u \models \psi$ for every closest $[\![\varphi]\!]$-world u to w.

Is it reasonable to assume that the comparative similarity relation is well-founded? Are there propositions $[\![\varphi]\!]$ such that for every $[\![\varphi]\!]$-world u some $[\![\varphi]\!]$world v exists that is more similar to w than u is — so that one can get closer and closer to w without ever getting in a $[\![\varphi]\!]$-world that is closest to w? It is not difficult to think of examples. How tall would you be in the closest world in which you are taller than you actually are?

The logic generated by the semantics sketched above is given by the following axioms and rules:

(Taut):	If φ has the form of a classical tautology, then $\vdash \varphi$
(MP$\rightarrow$):	$\varphi \rightarrow \psi, \varphi \vdash \psi$
(CI):	$\vdash \varphi \rightsquigarrow \varphi$
(CC):	$\vdash ((\varphi \rightsquigarrow \psi) \wedge (\varphi \rightsquigarrow \chi)) \rightarrow (\varphi \rightsquigarrow (\psi \wedge \chi))$
(CW):	$\vdash (\varphi \rightsquigarrow \psi) \rightarrow (\varphi \rightsquigarrow (\psi \vee \chi))$
(ASC):	$\vdash ((\varphi \rightsquigarrow \psi) \wedge (\varphi \rightsquigarrow \chi)) \rightarrow ((\varphi \wedge \psi) \rightsquigarrow \chi)$
(AD):	$\vdash ((\varphi \rightsquigarrow \chi) \wedge (\psi \rightsquigarrow \chi)) \rightarrow ((\varphi \vee \psi) \rightsquigarrow \chi)$
(REA):	If $\vdash \varphi \leftrightarrow \psi$, then $\vdash (\varphi \rightsquigarrow \chi) \leftrightarrow (\psi \rightsquigarrow \chi)$
(REC):	If $\vdash \varphi \leftrightarrow \psi$, then $\vdash (\chi \rightsquigarrow \varphi) \leftrightarrow (\chi \rightsquigarrow \psi)$

Here, (CI) is short for *Conditional Identity* , (CC) for *Conjunction of Consequents*, (CW) for *Weakening the Consequent*, (SAC) for *Strengthening the Antecedent with a Consequent*, (AD) for *Disjunction of Antecedents*, (MP$\rightarrow$)for *Modus Ponens for* $\rightarrow$, (REA) for *Replacement of Equivalent Antecedents*, and (REC) for *Replacement of Equivalent Consequents*.

This system, called **P**, is for conditional logic what **K** is for modal logic: it is the minimal system, which you get if you assume that the relations $\prec_w$ are just partial orderings[2] and have no additional properties. That **P** is (weakly)

1. Here '$u \preceq_w v$' is short for '$u = v$ or $u \prec_w v$'.
2. Actually, the only property that matters is transitivity. Irreflexivity is not expressible.

complete with respect to the class of partial orders was first proved by Burgess in [4]. This proof has been simplified by Friedman and Halpern in [12]. An altogether different proof of (strong) completeness is given in Veltman[23].

If in **P** the scheme (ASC) is strengthened to

$$\textit{Strengthening the Antecedent (AS)}: \quad (\varphi \rightsquigarrow \chi) \rightarrow ((\varphi \wedge \psi) \rightsquigarrow \chi)$$

one gets the system $\mathbf{K}^{\rightsquigarrow}$, which is just **K** in disguise.[3] More precisely,

Exercise 1

(i) In the language of modal logic, define $\varphi \rightsquigarrow \psi$ by $\Box(\varphi \rightarrow \psi)$. Suppose $\Delta \cup \{\varphi\}$ consists of formulas of the language of conditional logic. Then $\Delta \vdash_{\mathbf{K}^{\rightsquigarrow}} \varphi$ iff $\Delta \vdash_{\mathbf{K}} \varphi$.

(ii) In the language of conditional logic, define $\Box\varphi$ as $\neg\varphi \rightsquigarrow \bot$. Suppose $\Delta \cup \{\varphi\}$ consist of formulas of the language of modal logic. $\Delta \vdash_{\mathbf{K}} \varphi$ iff $\Delta \vdash_{\mathbf{K}^{\rightsquigarrow}} \varphi$.

It is straightforward[4] to prove (i) and (ii) from left to right. To prove (i) from right to left use (ii) from left to right, and similarly for (ii) from right to left use (i) from left to right.

Does the Limit Assumption make a difference to the logical properties of $\rightsquigarrow$? It does, but only for arguments with infinitely many premises. Under the Limit Assumption compactness fails.

Exercise 2

Let $p_1, \ldots, p_n, \ldots$ be countably many distinct atomic sentences, and let φ_k, and ψ_k, for any k be defined as follows:

$$\varphi_k = ((p_1 \vee \ldots, p_{k+1}) \rightsquigarrow \neg(p_1 \vee \ldots, p_k))$$

$$\psi_k = \neg((p_1 \vee \ldots, p_{k+1}) \rightsquigarrow (p_1 \vee \ldots, p_k))$$

Consider the set Δ consisting of all φ_k's and ψ_k's. The Limit Assumption holds iff Δ is not satisfiable.

3. Reminder: the minimal modal system **K** is given by the following axioms and rules:

(TAUT): If φ has the form of a classical tautology, then $\vdash \varphi$
(K-axiom): $\vdash \Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$
(MP$\rightarrow$): $\varphi \rightarrow \psi, \varphi \vdash \psi$
(NEC): If $\vdash \varphi$, then $\vdash \Box\varphi$

4. Straightforward but time consuming, even if one takes for granted that the *Deduction Theorem* and *Replacement of (Logical) Equivalents* hold both for **K** and $\mathbf{K}^{\rightsquigarrow}$.

3 Further Constraints

So far no constraints have been imposed on the comparative similarity relation $\prec$ that distinguish it from any other other relation that holds between three objects u, v and w when 'u is more ... to w than v. What extras does the fact one has to fill the dots with the word 'similar' bring?

Weak Centering: $w \in W_w$ for every $w \in W$, and for no $v \in W_w$ it holds that $v \prec_w w$.

Imposing this constraint means the next axiom gets valid. [5]

$$\textit{Modus Ponens for } \rightsquigarrow \text{ (MP}^{\rightsquigarrow}\text{)}: \quad (\varphi \rightsquigarrow \psi) \rightarrow (\varphi \rightarrow \psi)$$

Weak centering says that no world can be closer to a world w than w itself. If in addition you think that no world different from w can be equally close to w as w itself, you get this.

Strong Centering: $w \in W_w$ for every $w \in W$, and for every $v \in W_w$ such that $v \neq w$, $w \prec_w v$.

The logical pay off is this:

$$\textit{Conjunctive Sufficiency}: \quad (\varphi \wedge \psi) \rightarrow (\varphi \rightsquigarrow \psi)$$

If in establishing similarities and dissimilarities *all* characteristics of the worlds are taken into consideration, one of the consequences will be that only the world w itself will resemble the world w as much as the world w does. But in cases in which only *some* characteristics matter, there will often be more than one world that is just like w in all relevant respects. In these cases the structures will satisfy *Weak Centering*, but not *Strong Centering*

If you believe that two different worlds cannot be equally close to the actual one, you will support the following constraint:

Connectedness: for any $u, v \in W_w$, either, $u = w$, or $u \prec_w v$, or $v \prec_w u$.

In the presence of the Limit Assumption Connectedness implies that there will always be for any antecedent φ at most one $[\![\varphi]\!]$-world most resembling the actual world. This uniqueness assumption brings the following principle in its train:

$$\textit{Conditional Excluded Middle } \text{(CEM)}: \quad (\varphi \rightsquigarrow \psi) \vee (\varphi \rightsquigarrow \neg\psi)$$

Couldn't there be cases where we have several $[\![\varphi]\!]$-worlds, all equally close to the actual world and all closer to the actual world than any other world? In [16]

5. Reminder: A formula φ is *valid* on a frame $\langle W, \prec\rangle$ iff for every valuation V and every world $w \in W$ it holds that $\langle W, \prec, V\rangle, w \models \varphi$.
A formula φ is *valid* in a class $\mathfrak{C}$ of frames iff it is valid on all frames in $\mathfrak{C}$.

Lewis brings in the following example, due to W.V.O. Quine, to show that such cases do exist:

If Bizet and Verdi had been compatriots, Bizet would have been Italian. (e)

If Bizet and Verdi had been compatriots, Verdi would have been French. (f)

Now, if there is only one world closest to the actual world in which Bizet and Verdi are compatriots, it is impossible that both (e) and (f) are false while (g) is true:

> *If Bizet and Verdi had been compatriots, either Verdi would have been French or Bizet would have been Italian.* (g)

According to Lewis one can accept (g) without having to accept (e) or (f), and so he rejects the uniqueness assumption.

Lewis does accept the following constraint:

Almost-Connectedness: for any $u, v, w \in W_z$, if $u \prec_z w$, then either $u \prec_z v$ or $v \prec_z w$.

Define $u \simeq_w v$ iff neither $u \prec_w v$ nor $u \prec_w v$. The relation $\simeq_w$ is reflexive, and symmetric, but not necessarily transitive. Requiring that the relation $\prec_w$ is almost connected amounts to requiring that $\simeq_w$ is transitive.

Exercise 3 Check this.

In that case we can read '$u \simeq_w v$' as 'u and v are equally similar to w', and we can picture the relation $\prec_w$ as a linear order of equivalence classes of worlds. The corresponding axiom scheme is this:

Strengthening with a Possibility (ASP): $(\neg(\varphi \rightsquigarrow \neg\psi) \wedge (\varphi \rightsquigarrow \chi)) \rightarrow ((\varphi \wedge \psi) \rightsquigarrow \chi)$

The axiom ASP says that an antecedent of a counterfactual $\varphi \rightsquigarrow \chi$ may be strengthened with a formula ψ provided that the counterfactual assumption φ does not exclude the possibility that ψ. So, given the validity of ASC, this leaves only one case in which it is not allowed to strengthen the antecedent of a counterfactual $\varphi \rightsquigarrow \chi$ with the formula ψ. That's when $\varphi \rightsquigarrow \neg\psi$ is true and $\varphi \rightsquigarrow \psi$ is false. In the other three cases:

1. $\varphi \rightsquigarrow \psi$ is true, $\varphi \rightsquigarrow \neg\psi$ is true
2. $\varphi \rightsquigarrow \psi$ is true, $\varphi \rightsquigarrow \neg\psi$ is false
3. $\varphi \rightsquigarrow \psi$ is false, $\varphi \rightsquigarrow \neg\psi$ is false

strengthening the antecedent φ with ψ is valid.

Is it reasonable to assume that the comparative similarity relation is almost connected? Everybody who has tried to analyze the notion of comparative

similarity and to explain how it comes about, concluded that it is not.[6] Still, it is not easy to find a convincing counterexample to ASP. Ginsberg [10] suggests:

It's not the case that if Verdi and Satie had been compatriots, Satie and Bizet would not have been compatriots.
If Verdi and Satie had been compatriots, Bizet would have been French
If both Verdi and Satie, and Satie and Bizet had been compatriots, Bizet would have been French.

Despite this counterexample and the theoretical arguments underlying it, presently the most popular system for counterfactuals is given by $\mathbf{P}$ + ASP + MP$^{\rightsquigarrow}$.

Exercise 4
Check some of the correspondences mentioned in this section. More precisely, let $\mathfrak{F} = \langle W, \prec \rangle$ be any frame (with $\prec$ a strict partial order). Then the following holds

(i) MP$^{\rightsquigarrow}$ is valid on $\mathfrak{F}$ iff the ordering $\prec$ is weakly centered.
(ii) Both MP$^{\rightsquigarrow}$ and *Conjunctive Sufficiency* are valid on $\mathfrak{F}$ iff the ordering $\prec$ is strongly centered.
(iii) ASP is valid on on $\mathfrak{F}$ iff the ordering $\prec$ is almost connected.

Exercise 5
This exercise is about the following principle:

$$((\varphi \vee \psi) \rightsquigarrow \chi) \rightarrow ((\varphi \rightsquigarrow \chi) \wedge (\psi \rightsquigarrow \chi))$$

(i) Show that this principle is invalid in the system $\mathbf{P}$ + ASP + MP$^{\rightsquigarrow}$.
(ii) What do you think about the 'intuitive' plausibility of this principle?
(iii) Why not add this principle to $\mathbf{P}$ + ASP + MP$^{\rightsquigarrow}$?

4 Criticizing Comparative Similarity

Lewis's theory is still the most popular theory of conditionals around, despite the fact that right form the from the beginning philosophers and logicians have heavily criticized the ideas on which it is built.

The Tsjech logician Pavel Tichy came up with the following example:

> 'Consider a man, call him Jones, who is possessed of the following dispositions as regards wearing his hat. Bad weather invariably induces him to wear a hat. Fine weather, on the other hand, affects him neither way: on fine days he puts his hat on or leaves it on the peg, completely at random. Suppose moreover that actually the weather is bad, so Jones *is* wearing his hat.' [21]

6. For a critical analysis of the notion comparative similarity see Fine[6], Veltman[22], [25], Tichy[21], Pollock[18], Lewis[17], Kratzer[13], Kratzer[14].

The question is: would you accept the sentence *'If the weather had been fine, Jones would have been wearing his hat'*?[7]

Presumably, your answer is 'no', but Lewis's recipe would give 'yes'. In the actual world, it is raining and Jones is wearing is hat. Given that it is a matter of chance whether or not Jones wears his hat when the weather is fine, it would seem that for any sunny world in which Jones is not wearing his hat there is an equally sunny world in which he does, and which – because of this – is less different from the actual world.

Lewis is ready to admit that Tichy's example shows that the relevant conception of minimal difference needs to be spelled out with care, but he does not think the example shows that the idea of minimal difference is wrong. Perhaps such contingencies like whether or not Jones is wearing his hat, do not matter when the differences and similarities of possible worlds have to be assessed. This is at least what Lewis suggests in [17], where he formulates a system of weights that governs the notion of similarity involved. After some remarks on the important role of 'general' laws in this matter,[8] he says the following about the role of 'particular' fact.

> 'It is of little or no importance to secure approximate similarity of particular fact.' [17]

Here is a variant[9] of Tichy's puzzle which shows that this is not quite right.

> Suppose that Jones always flips a coin before he opens the curtains to see what the weather is like. Heads means he is going to wear his hat in case the weather is fine, whereas tails means he is not going to wear his hat in that case. Like above, bad weather invariably makes him wear his hat. Now suppose that today heads came up when he flipped the coin, and that it is raining. So, again, Jones is wearing his hat.

And again, the question is whether you would accept the sentence *'If the weather had been fine, Jones would have been wearing his hat'*. This time, your answer will be 'yes'. Lewis, too, would want to say 'yes', I guess. But can he? If similarity of particular fact did not matter in the first version of the puzzle, why would it now?

What really matters is this: In both cases Jones is wearing his hat *because* the weather is bad. In both cases we have to give up the proposition that the weather is bad — the very *reason* why Jones is wearing his hat. So, why should we want to keep assuming that he has his hat on? In the first case there is no

7. If you like the sentence better if there is a '*still*' between 'would' and 'have' in the consequent, then please read it that way.

8. As the first and the third criterion he mentions the following: *It is of the first importance to avoid big, widespread, diverse violations of law.... It is of the third importance to avoid even small, localized, simple violations of law.*

9. The example was suggested to me years ago by my former student Frank Mulkens.

special reason to do so; hence, we do not. In the second case there is a special reason. We will keep assuming that Jones is wearing his hat because we do not want to give up the independent information that the coin came down heads. And this, together with the counterfactual assumption that the weather is fine, brings in its train that Jones would have been wearing his hat.

In other words, similarity of particular fact is important, but only for facts that do not depend on other facts. Facts stand and fall together. In making a counterfactual assumption, we are prepared to give up everything that depends on something that we must give up to maintain consistency. But we want to keep in as many independent facts as we can. Later in this course we will develop this idea more precisely.

5 Non-monotonic consequence relations

The standard model theoretic notion of logical validity is monotonic: if ψ follows from $\varphi_1, \ldots, \varphi_n$, then ψ follows from $\varphi_1, \ldots, \varphi_n, \varphi_{n+1}$. This is so, because the standard notion requires that the conclusion be true in *any* model in which the premises are true, and, clearly, if ψ is true in *any* model in which $\varphi_1, \ldots, \varphi_n$ are true, then certainly so in *any* model in which $\varphi_1, \ldots, \varphi_n$ plus φ_{n+1} are true.

Non-monotonic logic started when in the late seventies logicians working in Artificial Intelligence noticed that in many practical situations when people draw a conclusion, they do not reckon with all conceivable possibilities left open by the premises, but only with some of these, the *most normal* ones or the ones *most likely* to occur. Something similar happened in the field of epistemic logic when at some point one got interested in arguments in which the premises represent 'all that is known'. In such cases the question is not so much whether the conclusion holds in all situations in which the premises hold, but whether it holds in the *'most ignorant'* situations among these.

There are more examples in which the phrase *'any model'* occurring in the definition of the standard notion of validity is restricted to 'the most ... models', where the dots are to be filled by some adjective. All these alternative notions of validity can be formally captured by assuming that the models of the language are ordered by a well-founded partial ordering $\prec$ and to stipulate that ψ is a (non-monotonic) consequence of $\varphi_1, \ldots, \varphi_n$ iff ψ is true in all models that are $\prec$-minimal in the class of models in which the premises $\varphi_1, \ldots, \varphi_n$ are true.

This must remind the reader of the frames and the truth-condition for counterfactuals introduced in the preceding section. Indeed, we are dealing here with a special case of the framework introduced there. In addition to the Limit Assumption, the following constraints are at stake.

Universality:	for every $w \in W, W_w = W$.
Absoluteness:	for every $u, w \in W, \prec_u = \prec_w$.

Absoluteness says that the relation $\prec_w$ is in fact independent of w, so that one can omit the subscript. *Universality* adds that $\prec$ is an ordering of the set of all possible worlds. So, the relations $\prec_w$ are all equal to one and the same well-founded partial ordering $\prec$ of the set of all possible worlds.

Secondly, given *Universality* and *Absoluteness*, if a sentence of the form $\varphi \rightsquigarrow \psi$ is true in one world of a model $\mathfrak{M}$, it will in fact be true in every world of $\mathfrak{M}$. This means that the following holds:

$$\mathfrak{M} \models \varphi \rightsquigarrow \psi \text{ iff } \mathfrak{M}, w \models \psi \text{ for every } \prec\text{-minimal world } w \text{ in } [\![\varphi]\!].$$

Finally, let's write '$\varphi_1, \ldots, \varphi_n \mid\!\sim \psi$' instead of '$(\varphi_1 \wedge \ldots \wedge \varphi_n) \rightsquigarrow \psi$', and '$\varphi_1, \ldots, \varphi_n \mid\!\sim_{\mathfrak{M}} \psi$' instead of '$\mathfrak{M} \models (\varphi_1 \wedge \ldots \wedge \varphi_n) \rightsquigarrow \chi$'. In doing so, we arrive at what in Kraus et al.[15] appears as the definition of 'the entailment relation $\mid\!\sim_{\mathfrak{M}}$ defined by the model $\mathfrak{M}$'.

(*) $\varphi_1, \ldots, \varphi_n \mid\!\sim_{\mathcal{M}} \psi$ iff $\mathfrak{M}, w \models \psi$ for every $\prec$-minimal world w in $[\![\varphi_1]\!] \cap \ldots \cap [\![\varphi_n]\!]$.

The authors of [15] refer to the relation $\prec$ as a preference relation, and to the models $\mathfrak{M} = \langle \mathfrak{W}, \prec, \mathfrak{V} \rangle$ as preferential models. They are interested in the properties of the *preferential* consequence relation $\mid\!\sim$, formally modeled by (*).

It will come as no surprise that $\mid\!\sim$ behaves like a counterfactual implication $\rightsquigarrow$. However, there is an important syntactic difference between between $\mid\!\sim$ and $\rightsquigarrow$. Conditionals sometimes occur nested in other conditionals — as in $\varphi \rightsquigarrow (\psi \rightsquigarrow \chi)$ — but nesting sentences expressing an entailment relation is quite incomprehensible. The entailment relation belongs to the metalanguage rather than the object language. What could $\varphi \mid\!\sim (\psi \mid\!\sim \chi)$ possibly mean?

This, however, does not give rise to important semantic differences between $\mid\!\sim$ and $\rightsquigarrow$.

Proposition 1
Let Δ be a set of formulas containing only non-nested conditionals. If Δ is satisfiable on any frame, then it is satisfiable on a frame with a universal and absolute $\prec$ relation.[10]

Given this, one might expect the system **P** to give a complete characterization of the properties of $\mid\!\sim$. Kraus et al.[15], using the methods of [23], prove that this is indeed the case. One easily recognizes the axiom schemes introduced

10. This proposition does not hold for arbitrary sets of formulas. If nesting is allowed one has to add the **S5** axioms $\Box\varphi \rightarrow \varphi$, $\Box\varphi \rightarrow \Box\Box\varphi$, and $\Diamond\varphi \rightarrow \Box\Diamond\varphi$ to **P** in order to get a system that is complete with respect to the universal and absolute frames. (Here $\Box\varphi =_{df} \neg\varphi \rightsquigarrow \bot$.)

in the previous section in the next principles of entailment.

$$\begin{array}{rl}
\text{(CI) becomes } \textit{Reflexivity}: & \varphi \mathrel{|\!\sim} \varphi \\
\text{(CC) becomes } \textit{And}: & \text{If } \varphi \mathrel{|\!\sim} \psi \text{ and } \varphi \mathrel{|\!\sim} \chi, \text{ then } \varphi \mathrel{|\!\sim} (\psi \wedge \chi) \\
\text{(CW) becomes } \textit{Right Weakening}^{11}: & \text{If } \varphi \mathrel{|\!\sim} \psi \text{ and } \psi \models \chi, \text{ then } \varphi \mathrel{|\!\sim} \chi \\
\text{(ASC) becomes } \textit{Cautious Monotony}: & \text{If } \varphi \mathrel{|\!\sim} \psi \text{ and } \varphi \mathrel{|\!\sim} \chi, \text{ then } \varphi, \psi \mathrel{|\!\sim} \chi \\
\text{(AD) becomes } \textit{Or}: & \text{If } \varphi \mathrel{|\!\sim} \chi \text{ and } \psi \mathrel{|\!\sim} \chi, \text{ then } (\varphi \vee \psi) \mathrel{|\!\sim} \chi \\
\text{(REA) becomes } \textit{Left Logical Equivalence}: & \text{If } \varphi \models \psi \text{ and } \psi \models \varphi, \text{ then} \\
 & \quad \text{if } \varphi \mathrel{|\!\sim} \chi, \text{ also } \psi \mathrel{|\!\sim} \chi
\end{array}$$

The literal translation of (CW) would be ' If $\varphi \mathrel{|\!\sim} \psi$, then $\varphi \mathrel{|\!\sim} \psi \vee \chi$'. *Right Weakening* is equivalent to this. ($\models$ stands for the classical entailment relation.)

As a characterization of an entailment relation the system **P** is a bit odd. One would expect only purely structural principles. The principles *Or*, and *And*, however, presuppose that the object language has connectives with the properties of conjunction and disjunction. Kraus et al.[15] also discuss a weaker system consisting of only structural rules. It is called **C**, where '**C**' stands for 'cumulative', and it was originally proposed by Dov Gabbay[7] as a system describing the weakest reasonable consequence relation. It is given by: *Reflexivity, Right Weakening, Cautious Monotony, Left Logical Equivalence*, and

$$\textit{Cut}: \quad \text{If } \varphi, \psi \mathrel{|\!\sim} \chi \text{ and } \varphi \mathrel{|\!\sim} \psi, \text{ then } \varphi \mathrel{|\!\sim} \chi$$

It is left to the reader to show that *Cut* is a derived rule of **P**.

An important field in which a non-monotonic consequence relation is employed is the field of default reasoning. Actually, in the modal approach to default reasoning not only the consequence relation but also the defaults rules themselves are modeled after conditionals. Read '$\varphi \rightsquigarrow \psi$' as '*If φ, then normally ψ*', and take the underlying well-founded ordering $\prec$ of the set of possible worlds to be the relation '*... is more normal than...*'. Then a rule $\varphi \rightsquigarrow \psi$ will hold in a model if ψ is true at the most normal $[\![\varphi]\!]$-worlds. An agent who has learnt that φ is the case and who accepts the rule $\varphi \rightsquigarrow \psi$ will expect that ψ is the case provided there is no evidence that the case at hand is exceptional.

More generally, default rules are of crucial importance when some decision must be made in circumstances where the facts of the matter are only partly known. In such a case one must reckon with several possibilities. Default rules serve to narrow down this range of possibilities: some of these possibilities are more normal than other. An agent will expect that the actual world conforms to as many standards of normality as possible given the information at hand.

Several theories have been developed that formalize this phenomenon. They differ in the way they formally capture the idea that an agent will expect the actual world to be as normal as possible given the circumstances described by the premises. James Delgrande [5] was the first who proposed a definition

for the set of worlds that best meet the agent's expectations. Alternative definitions are proposed in Asher & Morreau[2] and Veltman [24]. See [3] for a detailed comparison of these theories and Halpern et.al.[12] for technical insights.

6 Belief revision

There is still another way to read $\varphi \rightsquigarrow \psi$: 'After a revision by φ, it is believed that ψ'. Here the topic is belief revision, and the question at stake is how an agent should change his or her beliefs in the face of new information. The formula φ is supposed to bring new information — possibly contradicting the information available — and if $\varphi \rightsquigarrow \psi$ is true, this means that ψ is accepted after the incorporation φ in ones stock of beliefs.

Checking the axioms for $\rightsquigarrow$ with this reading in mind, we find that many of them sound quite plausible. For example: *Conditional Identity*, $\varphi \rightsquigarrow \varphi$, becomes *'After a revision by φ, it is believed that φ'*, and *Disjunction of Antecedents*, $(\varphi \rightsquigarrow \chi) \wedge (\psi \rightsquigarrow \chi)) \rightarrow ((\varphi \vee \psi) \rightsquigarrow \chi)$, can be read as '*If both a revision with φ and a revision by ψ lead to the belief χ, then so does a revision by $\varphi \vee \psi$*'. Are we here once more dealing with **P** or one of its extensions?

Let's start at the beginning. In 1985 Carlos Alchourron, Peter Gärdenfors and David Makinson published a by now classic paper [1] in which they discuss three forms of belief change: expansion, contraction and revision. Modeling an agents beliefs by a deductively closed theory K, called a belief set, a number of rationality postulates are laid down for the expansion K^+_φ of K by φ, the contraction K^-_φ of K by φ, and the revision K^*_φ of K by φ.

The constraints for expansion uniquely determine K^+_φ as the set $\{\psi \mid K, \varphi \vdash \psi\}$. The constraints for contraction and revision do not uniquely determine K^-_φ and K^*_φ because the outcomes of these operations do not depend on logical factors only. Epistemic factors may also play a role. For example, in revising their beliefs agents may be prepared to give up one sentence rather than the other because the empirical support for the one is much better than for the other.

Here are the so-called AGM postulates for revision as formulated in Gärdenfors[8]:

K^{*1} For any sentence φ and any belief set K, K^*_φ is a belief set
K^{*2} $\varphi \in K^*_\varphi$
K^{*3} $K^*_\varphi \subseteq K^+_\varphi$
K^{*4} If $\neg\varphi \notin K$, then $K^+_\varphi \subseteq K^*_\varphi$
K^{*5} $K^*_\varphi = \{\psi \mid \bot \vdash \psi\}$ iff $\vdash \neg\varphi$;
K^{*6} If $\vdash \varphi \leftrightarrow \psi$, then $K^*_\varphi = K^*_\psi$
K^{*7} $K^*_{\varphi\wedge\psi} \subseteq (K^*_\varphi)^+_\psi$
K^{*8} If $\neg\varphi \notin K$, then $(K^*_\psi)^+_\varphi \subseteq K^*_{\varphi\wedge\psi}$

Adam Grove[11] was the first to notice that the semantics for counterfactuals as defined in section 1, supplies an interpretation for these postulates. For every

belief set K, we consider the set of *models for* K, where a model $\mathfrak{M}_{\mathfrak{K}}$ for K is given by $\mathfrak{M}_{\mathfrak{K}} = \langle \mathfrak{W}, \prec, \mathfrak{V} \rangle$, where

- W is the set of all maximal consistent theories of the language in which K is formulated;
- $\prec$ is a well-founded and almost connected strict partial ordering of W such that the $\prec$-minimal elements of W are given by the set of maximal consistent extensions of K;
- $V(p)(w) = 1$ iff $p \in w$.

Proposition 2
Let $\mathfrak{M}_{\mathfrak{K}} = \langle \mathfrak{W}, \prec, \mathfrak{V} \rangle$ be a model for K. Define K^*_φ for every φ as follows:

$$\psi \in K^*_\varphi \text{ iff } \psi \in w \text{ for every } w \text{ such that } w \text{ is } \prec\text{-minimal in } [\![\varphi]\!].$$

Then the postulates $\mathbf{K^*1}$ to $\mathbf{K^*8}$ are satisfied.

Conversely, we have

Proposition 3
Let K^* be a revision function for some belief set K satisfying $\mathbf{K^*1}$ to $\mathbf{K^*8}$.
Define $\mathfrak{M}_{\mathfrak{K}} = \langle \mathfrak{W}, \prec, \mathfrak{V} \rangle$ as follows:

- W is the set of all maximal consistent theories of the language in which K is formulated;
- $u \prec w$ iff $\tau(w) \subseteq \tau(u)$ and $u \notin \tau(w)$.
  Here, τ is given by: $v \in \tau(w)$ iff $v \in W$ and there is some φ such that $K^*_\varphi \subseteq w$ and $\varphi \in v$.
- $V(p, w) = 1$ iff $p \in w$.

Then $\mathfrak{M}_{\mathfrak{K}} = \langle \mathfrak{W}, \prec, \mathfrak{V} \rangle$ is a model for K for which the following holds:

$$\psi \in K^*_\varphi \text{ iff } \psi \in w \text{ for every } w \text{ such that } w \text{ is } \prec\text{-minimal in } [\![\varphi]\!].$$

This means that whenever $\psi \in K^*_\varphi$, the model $\mathfrak{M}_{\mathfrak{K}}$ verifies $\varphi \rightsquigarrow \psi$. This model is almost connected. Therefore, in view the observations we made in section 1 and 5, it follows that the AGM revision constraints endow $\rightsquigarrow$ with the logic **P** + ASP.[12]

One may be tempted to conclude from the above that revising ones beliefs by φ and making the counterfactual assumption *if it had been the case that* φ amount to the same thing. However, even though these cognitive operations have much in common formally, there are huge differences between them. When you believe that φ is true and you try to imagine what would have been the case if φ

12. An extensive discussion of the AGM theory of belief revision and its representation in conditional logic can be found in [9].

had been false, you have to change your cognitive state, but it is it not the kind of change you would have to make if you were to discover that φ is *in fact* false. It is not a *correction*. Consider for example φ = *Oswald killed Kennedy*. Supposing that Oswald had not killed Kennedy might make you think 'If Oswald had not killed Kennedy, Kennedy might still be alive'. If, however, at some point you were to find out that your belief that Oswald killed Kennedy is in fact wrong, and you had to revise your beliefs accordingly, it is very likely that after this revision you would still believe that Kennedy is dead.[13]

7 Conditional Obligation

IN this section we will read $\varphi \rightsquigarrow \varphi$, as *'Given that φ, it is obligatory that ψ'*, or as *'Given that φ, it ought to be that ψ'*. Think of the underlying relation not as comparative similarity but of comparative 'goodness'. In other words, read '$u \prec_w v$' as *'In world w the world u is preferable to the world v'*. The source of this preference ordering my vary. As Lewis puts it:

> 'Perhaps the worlds are ordered according to their total net content of pleasure, measured by some hedonic calculus; or their content of beauty, truth and love; or their content of some simple non-natural quality. Perhaps they are ordered according to the extent that their inhabitants obey the law of God, of Nature, or of Man. Perhaps according to how well they measure up to some sort of standard of objective morality, if such there be; perhaps according to someone's personal taste in possible worlds; perhaps . . . It does not matter. We can build in the same way on any of these foundations or on others.' [16].

Exercise 6 Find an example which shows that AS should not hold for conditional obligation.

Exercise 7

(i) Which of the properties discussed in section 3 do you think are plausible for $\prec$ in this context?

(ii) And how about *Universality* and *Absoluteness*?

Given *Universality* and *Absoluteness*, if a sentence of the form $\varphi \rightsquigarrow \psi$ is true in one world of a model $\mathfrak{M}$, it will in fact be true in every world of $\mathfrak{M}$.

Absolute obligations can be defined in terms of conditional obligations. Define $O\varphi =_{df} \top \rightsquigarrow \varphi$ and read '$O\varphi$' as *'It ought to be the case that φ'*. Historically, absolute obligations were studied first, and only when it turned out that formulas of the form $O(\varphi \rightarrow \psi)$ could not play the role of conditional obligations

13. See [19] for an insightful discussion of these points.

systems like the one described here were introduced. It is common to abbreviate formulas of the form $\neg O \neg \varphi$ as $P\varphi$ and read the latter as *'It is permitted that φ'*.

Exercise 8

(i) Can you formulate a constraint (to be imposed on $\prec$) which ensures that $O\varphi \rightarrow P\varphi$ gets valid?

(ii) In none of the standard systems of deontic logic we have:
$P(p \vee q) \models Pp$.
Yet, intuitively, *You may take an apple or a pear* implies
You may take an apple.
Think about this.

References

[1] C. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction functions and their associated revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[2] N. Asher. Commonsense entailment: A conditional logic for some generics. In G. Crocco, L. Fari nas del Cerro, and A. Herzig, editors, *Conditionals: From Philosophy to Computer Science*, pages 103–145. Oxford University Press, Oxford, 1995.

[3] R. Blutner. *Normality* in update semantics. In M. Simons and T. Galloway, editors, *Proceedings from Semantics and Linguistic Theory V*, pages 19–36, Ithaca, New York, 1995. Cornell University.

[4] J. P. Burgess. Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic*, 22(1):76–84, 1981.

[5] J. P. Delgrande. Conditional logics for defeasible logics. In Phillipe Besnard and Anthony Hunter, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 2: Reasoning with Actual and Potential Contradictions*, pages 135–174. Kluwer Academic Publishers, Dordrecht, 1998.

[6] K. Fine. Critical review of D. Lewis's 'Counterfactuals'. *Mind*, 84:451–458, 1975.

[7] D. Gabbay. Theoretical foundations for non-monotonic reasoning in expert systems. In K. Apt, editor, *Proceedings of the NATO Advanced Study Institute on Logics and Models of Concurrent Systems*, pages 439–457. Springer-Verlag, 1985.

[8] P. Gärdenfors. *Knowledge in Flux*. The MIT Press, Cambridge, Massachusetts, 1988.

[9] P. Gärdenfors and H. Rott. Belief revision. In J.A. Robinson D. Gabbay, C.J. Hogger, editor, *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 4*, pages 35–132. Oxford UNiversity Press, 1996.

[10] M.L. Ginsberg. Counterfactuals. *Artificial Intelligence*, 30:35–79, 1986.
[11] A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.
[12] J. Halpern, N. Friedman, and D. Koller. First-order conditional logic for default reasoning revisited. *ACM Transactions on Programming Languages and Systems*, 1(2):175–207, 2000.
[13] A. Kratzer. Partition and revision: the semantics of counterfactuals. *Journal of Philosophical Logic*, 10:242–258, 1981.
[14] A. Kratzer. An investigation of the lumps of thought. *Linguistics and Philosophy*, 87(1):3–27, 1989.
[15] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 14(1):167–207, 1990.
[16] D. Lewis. *Counterfactuals*. Basil Blackwell, Oxford, 1973.
[17] D. Lewis. Counterfactual dependence and time's arrow. *Noûs*, 13:455–476, 1979.
[18] J. Pollock. *Subjunctive Reasoning*. Reidel, Dordrecht, 1976.
[19] H. Rott. Moody conditionals: Hamburgers, switches, and the tragic death of an american president. In Jelle Gerbrandy, Maarten Marx, Maarten de Rijke, and Yde Venema, editors, *JFAK. Essays dedicated to Johan van Benthem on the occasion of his 50th birthday*, pages 98–112. Amsterdam University Press, Amsterdam, 1999.
[20] R Stalnaker. A theory of conditionals. In Nicholas Rescher, editor, *Studies in Logical Theory*, pages 98–112. Basil Blackwell, Oxford, 1968.
[21] P. Tichy. A counterexample to the stalnaker-lewis analysis of counterfactuals. *Philosophical Studies*, 29:271–273, 1976.
[22] F. Veltman. Prejudices, presuppositions, and the theory of counterfactuals. In J. Groenendijk and M. Stokhof, editors, *Amsterdam Papers in Formal Grammar. Proceedings of the 1st Amsterdam Colloquium*, pages 248–281. University of Amsterdam, 1976.
[23] F. Veltman. *Logics for Conditionals*. Ph.D. dissertation, University of Amsterdam, Amsterdam, 1985.
[24] F. Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25:221–261, 1996.
[25] F. Veltman. Making counterfactual assumptions. *Journal of Semantics*, 22:159–180, 2005.