Le conditionnel de Lewis

M. Cozic (DEC, ENS) & P. Égré (CNRS, IJN)

Textes de référence :

- D.K. Lewis (1973a) *Counterfactuals*, Harvard UP, Cambridge (Mass.), 1973 [L'exposé complet de la théorie de Lewis : ses motivations conceptuelles, ses fondements philosophiques et son étude logique approfondie]

- D.K. Lewis (1973b) "Counterfactuals and Comparative Possibility", *Journal of Philosophical Logic*, 2, 1973, pp. 418-46 [Précis de *Counterfactuals*.]

1 Les présupposés de la théorie de Stalnaker

• Lewis rejette l'analyse des conditionnels en termes de conditionnels stricts. Dans (1973b), il donne deux familles de contre-exemples :

1. $(\phi \hookrightarrow \neg\psi) \models ((\phi \wedge \psi) \hookrightarrow \neg\chi) \wedge ((\phi \wedge \psi) \hookrightarrow \chi)$

   (1) If Albert had come to the party, he would not have brought Betty.

   (2) I Albert had come to the party and had brought Betty, Carl would not have stayed.

2. $(\phi \hookrightarrow \psi) \models ((\phi \wedge \chi) \hookrightarrow \psi)$ (renforcement)

   (3) If I had shirked my duty, no harm would have ensued.

   (4) If I had shirked my duty and you had too, harm would have ensued.

• Lewis accepte les intuitions fondamentales de la théorie de Stalnaker (1968) - qu'il appelle ANALYSE 1 - : les mondes possibles et les relations de similitude entre mondes. Il y a toute fois deux présupposés dans la théorie de Stalnaker qu'il explicite *et rejette*.

1.1 L'hypothèse d'unicité (*Uniqueness Assumption*)

• La fonction de sélection de Stalnaker est une fonction à valeur dans W ; cela signifie qu'à toute paire constituée d'une formule ϕ et d'un monde d'évaluation w, elle associe un et un seul monde possible : le monde $f([[\phi]], w)$. Aussi, il existe un unique monde qui est le plus proche de w et où ϕ est vraie.

• Lewis rejette cette hypothèse : pourquoi n'existerait-il pas plusieurs mondes possibles qui seraient tous "aussi similaires" au monde d'évaluation w et où ϕ serait vraie ?

Exemple :

Lewis (à la suite de Quine, *Methods of Logic*, qui en faisait un argument contre la possibilité d'une analyse cohérente des contrefactuels !) propose la paire suivante de conditionnels :

(5) Si Bizet et Verdi étaient compatriotes (A), ils seraient italiens (I)

(6) Si Bizet et Verdi étaient compatriotes (A), ils seraient français (F)

Si l'on accepte la théorie de Stalnaker, alors il existe un et un seul monde qui est le plus proche du monde actuel ; par conséquent, l'un des deux conséquents seulement est vrai, selon que le monde le plus proche est un I-monde ou un F-monde. Ce que soutient Lewis, c'est que dans des situations de ce genre, il paraît *ad hoc* de stipuler qu'il y aura forcément un monde plus proche que l'autre.

• Si l'on est convaincu par l'objection de Lewis à l'hypothèse d'unicité, on peut facilement amender la théorie de Stalnaker pour prendre en compte l'objection : il suffit de considérer les fonctions de sélection comme des *correspondances*, c'est-à-dire comme des fonctions f : $\wp(W) \times w \to \wp(W)$. La fonction de sélection ne sélectionne plus alors nécessairement un seul monde, mais éventuellement plusieurs.

Il faut bien sûr modifier les clauses modales de la relation de satisfaction. Au lieu d'avoir pour clause de satisfaction

$M, w \models (\phi \Box\!\!\rightarrow \psi)$ ssi $M, f([[\phi]], w) \models \psi$

on aurait alors quelque chose comme

$M, w \models (\phi \Box\!\!\rightarrow \psi)$ ssi pour tout $w' \in f([[\phi]], w)$, $M, w' \models \psi$

Autrement dit : $(\phi \Box\!\!\rightarrow \psi)$ est vraie en un monde w ssi ψ est vraie dans tous les ϕ-mondes accessibles les plus proches de w, s'il en existe. C'est ce que Lewis appelle l'ANALYSE 2. Dans cette proposition, le conditionnel se comporte comme une sorte d'opérateur de nécessité $\Box$ indexé par l'antécédent : $\Box_\phi$ pour ainsi dire.

Exemple :

Reprenons l'exemple de Bizet et de Verdi. Supposons que le monde actuel (où ils ne sont pas compatriotes) soit w, le monde le plus proche où ils sont français tous les deux soit w_f et le monde le plus proche où ils sont italiens tous les deux soit w_i. On a donc :

$M, w_f \models F$ mais $M, w_f \not\models I$
$M, w_i \models I$ mais $M, w_i \not\models F$

Puisque $f([[\phi]], w) = \{w_i, w_f\}$, il est clair que l'on a $M, w \not\models (A \Box\!\!\rightarrow F)$; on a également $M, w \not\models (A \Box\!\!\rightarrow \neg F)$. Par conséquent, on a aussi :

$M, w \not\models ((A \Box\!\!\rightarrow F) \vee (A \Box\!\!\rightarrow \neg F))$

Par conséquent *sans l'hypothèse d'unicité, le tiers-exclu conditionnel n'est plus valide.* C'est une conséquence intuitive si l'on songe à l'exemple de Bizet Verdi : si Bizet et Verdi avaient été compatriotes, il n'est pas sûr qu'ils auraient été français, mais il n'est pas sûr non plus qu'il n'auraient pas été français !

1.2 L'hypothèse de limite (***Limit Assumption***)

• Supposons que la relation de similitude entre mondes soit représentée par un ordre $\leq_w$: $u \leq_w v$ signifie que u est plus proche de w que ne l'est v. L'hypothèse d'unicité consiste à supposer que pour toute formule ϕ et tout monde w, il y a un seul ϕ-monde qui est $\leq_w$-minimal. L'affaiblissement que l'on a vu consiste à supposer qu'il peut exister plusieurs ϕ-mondes $\leq_w$-minimaux. Mais on suppose toujours qu'il existe des mondes $\leq_w$-minimaux. C'est ce que Lewis appelle l'hypothèse de limite : il existe une limite quand on "descend" une relation de similitude $\leq_w$ vers le monde w.

• Considérons le conditionnel (schématique) suivant :

(7) Si Paul mesurait plus de deux mètres, alors C.

Pour l'évaluer conformément aux théories précédentes (que ce soit celle de la fonction de sélection ou celle de la "correspondance" de sélection), il faut considérer les mondes possibles les plus proches où Paul mesure plus de deux mètres. Problème : on a une chaîne descendante infinie de mondes possibles. Si l'on prend un monde contrefactuel au hasard, où Paul mesure deux mètres et ϵ cm, alors on peut trouver un monde plus proche où Paul mesure deux mètres et $\epsilon/2$, etc.

• L'hypothèse de limite est ancrée dans les "correspondances" de sélection, de même que l'hypothèse d'unicité est ancrée dans les fonctions de sélection : si l'on rejette l'hypothèse, alors il faut rejeter également les correspondances de sélection et trouver une autre analyse sémantique.

• La théorie finale de Lewis cherche à prendre en compte les violations de l'hypothèse de limite. La modification qu'il apporte est simple : puisqu'il peut y avoir des chaînes descendantes de mondes de plus en plus proches du monde d'évaluation w, ce que l'on peut exiger (au lieu d'exiger que les ϕ-mondes possibles les plus proches soient des ψ-mondes) c'est qu'il existe un monde possible v qui soit à la fois un ϕ-monde et un ψ-monde tel qu'aucun monde u plus proche de w ne soit un ϕ-monde et un $\neg\psi$-monde. C'est ce que Lewis appelle l'ANALYSE 3 :

$(\phi \mathrel{\Box\!\!\rightarrow} \psi)$ est vraie en w ssi il existe un $(\phi \wedge \psi)$-monde accessible qui soit plus proche de w que tout $(\phi \wedge \neg\psi)$-monde accessible, s'il existe un quelconque ϕ-monde accessible.

2 Les modèles de similitude

• La sémantique de Lewis repose sur des relations de similitudes indexées par les mondes possibles ; ces relations n'obéissent pas nécessairement aux hypothèses d'unicité et de limite. En revanche, elles obéissent aux deux hypothèses suivantes :

1. *hypothèse d'ordre :* les $\leq_w$ sont des relations d'ordre faible : (1) totale : si v et u sont accessibles depuis w, alors $u \leq_w v$ ou $v \leq_w u$; (2) transitive : si $v \leq_w u$ et $u \leq_w r$, alors $v \leq_w r$.

2. *hypothèse de centrage :* pour tout monde w, w est accessible depuis lui-même (wRw) et est le monde le plus proche de lui-même (si $v \leq_w w$, alors $v = w$).

Définition 1
*Un **modèle de similitude** $M = \langle W, R, I, \{\leq_w\}_{w \in W}\rangle$ pour $\mathcal{L}_{\Box\!\rightarrow}$ est un n-uplet constitué*

1. *d'un ensemble W de mondes possibles*
2. *d'une relation d'accessibilité $R \subseteq W \times W$ réflexive*
3. *d'une valuation atomique $I : At \rightarrow \wp(W)$*
4. *d'une famille de relations de similitude $\{\leq_w\}_{w \in W}$ telle que*

   (i) $\leq_w$ constitue un préordre total i.e. $\leq_w$ est (1) totale : pour tous mondes u, v, $u \leq_w v$ ou $v \leq_w u$; (2) transitive : si $v \leq_w u$ et $u \leq_w r$, alors $v \leq_w r$.

   (ii) w est le monde le plus semblable à lui-même (si $v \leq_w w$, alors $v = w$)

   (iii) les mondes inaccessibles depuis w sont les mondes les moins semblables à w (si $\neg wRv$, alors pour tout u, $u \leq_w v$) et sont strictement moins semblables que les mondes accessibles (si $\neg wRv$ et wRu, alors $u <_w v$)

• Une fois les modèles de similitude définis, on peut formuler rigoureusement l'analyse sémantique de Lewis :

Définition 2 (Relation de satisfaction)
Soit $M = \langle W, R, I, \{\leq_w\}_{w \in W}\rangle$ un modèle de similitude ; pour tout $w \in W$, on définit ainsi la relation de satisfaction :

(i) $M, w \models p$ *ssi* $w \in I(p)$

(ii) $M, w \models \neg\phi$ *ssi* $M, w \not\models \phi$

(iii) $M, w \models (\phi \wedge \psi)$ *ssi* $M, w \models \phi$ *et* $M, w \models \psi$
$M, w \models (\phi \vee \psi)$ *ssi* $M, w \models \phi$ *ou* $M, w \models \psi$
$M, w \models (\phi \rightarrow \psi)$ *ssi si* $M, w \models \phi$ *alors* $M, w \models \psi$

(iv) $M, w \models (\phi \;\Box\!\!\rightarrow \psi)$ *ssi si* $[[\phi]] \cap R(w) \neq \emptyset$*, alors il existe un* $v \in R(w) \cap [[(\phi \wedge \psi)]]$ *tel qu'il n'existe aucun* u *tel que* $u \leq_w v$ *et* $u \in [[(\phi \wedge \neg\psi)]]$

Définition 3
Une formule ϕ est valide dans les modèles de similitude ssi elle est vraie en tout monde de tout modèle de similitude ; on le note

$$\boxed{\models_{sim} \phi}$$

3 Possibilité comparative

• La clause relative à " $\Box\!\!\rightarrow$" suggère qu'on peut définir le conditionnel en termes d'autres opérateurs modaux :

(i) d'une part, l'opérateur de possibilité : "si $[[\phi]] \cap R(w) \neq \emptyset$" veut simplement dire que $\Diamond\phi$ est vraie en w

(ii) d'autre part, d'un opérateur de *possibilité comparative* : "alors il existe un $v \in R(w) \cap [[(\phi \wedge \psi)]]$ tel qu'il n'existe aucun u tel que $u \leq_w v$ et $u \in [[(\phi \wedge \neg\psi)]]$" signifie en quelque sorte que $(\phi \wedge \psi)$ est "plus possible" que $(\phi \wedge \neg\psi)$

• De manière générale, on peut introduire l'opérateur modal binaire "$\prec$" de possibilité comparative dont la clause de satisfaction est la suivante :

$M, w \models (\phi \prec \psi)$ ssi il existe un $v \in R(w) \cap [[\phi]]$ tel qu'il n'existe aucun u tel que $u \leq_w v$ et $u \in [[\psi]]$

On peut alors effectivement définir " $\Box\!\!\rightarrow$" en termes de "$\Diamond$" et de "$\prec$" :

$(\phi \Box\!\!\rightarrow \psi) =_{df} (\Diamond\phi \rightarrow ((\phi \wedge \psi) \prec (\phi \wedge \neg\psi))$

• Remarques :

- notons que l'on peut également l'opérateur de possibilité en termes de possibilité comparative : si l'on stipule que $\perp$ n'est vraie dans aucun monde possible, alors on peut définir $\Diamond\phi$ comme $\phi \prec \perp$. Si $\phi \prec \perp$ en w, alors ϕ est vraie dans un monde accessible depuis w, et réciproquement.

- notons aussi que l'on peut définir l'opérateur de possibilité "$\Diamond$" et l'opérateur de possibilité comparative "$\prec$" à partir du conditionnel " $\Box\!\!\rightarrow$".

4 Systèmes de sphères

• Lewis (1973a) donne une seconde sémantique pour le conditionnel, basée sur la notion de sphère. Une sphère S (relativement à un monde w) est l'ensemble des mondes qui ont au moins un certain degré de similitude par rapport à w : S est une sphère (relativement à w) si tous les éléments de S sont accessibles depuis w et sont plus similaires à w que n'importe quel monde de S^c (les mondes qui ne sont pas dans la sphère S).

On peut également penser aux sphères en termes de relation de similitude : si v est dans une sphère relativement à w, alors tout monde plus semblable à w que v est également dans cette sphère : si $v \in S$, alors tout u t.q. $u \leq_w v$, $u \in S$. Une sphère S pour w est un ensemble de mondes clos par la relation $\leq_w$ (dans le sens descendant).

• Voici les hypothèses que Lewis propose sur les sphères:

1. $\mathcal{S}_w$ est centré sur w : $\{w\} \in \mathcal{S}_w$
2. $\mathcal{S}_w$ est emboîté : si $S, T \in \mathcal{S}_w$, alors $S \subseteq T$ ou $T \subseteq S$
3. $\mathcal{S}_w$ est clos sous l'union : si $\mathcal{S} \subseteq \mathcal{S}_w$, alors $\bigcup \mathcal{S} \in \mathcal{S}_w$
4. $\mathcal{S}_w$ est clos sous l'intersection (non vide) : si $\mathcal{S} \neq \emptyset \subseteq \mathcal{S}_w$, alors $\bigcap \mathcal{S} \in \mathcal{S}_w$

Commentaires: on peut justifier chacune des conditions en termes de similitudes entre mondes possibles en ayant en tête l'interprétation selon laquelle pour toute sphère S, si $u \in S$ mais $v \notin S$, cela signifie que u est strictement plus semblable à w que ne l'est v :

1. traduit l'hypothèse de centrage : le monde le plus semblable à w est w lui-même.

2. sinon, il existe deux mondes v et u tel que $v \in S$ mais $v \notin T$ et $u \in T$ mais $u \notin S$. Cela signifierait que u est strictement plus semblable à w que u et que u est strictement plus semblable à w que v.

3. supposons que $v \in \bigcup \mathcal{S}$ mais $u \notin \bigcup \mathcal{S}$. Cela signifie que v est dans l'une des sphères de $\mathcal{S}_w$ et que ce n'est pas le cas pour u. Donc v est plus sembable à w que u ne l'est. Si l'on se laisse guider par l'interprétation des sphères, cela légitime l'appartenance de $\bigcup \mathcal{S}$ à $\mathcal{S}_w$ - tout monde de $\bigcup \mathcal{S}$ est plus semblable à w que tout monde qui n'appartient pas à $\bigcup \mathcal{S}$.

3. même raisonnement.

Définition 4
Un ***modèle de sphères*** $M = \langle W, I, (\mathcal{S}_w)_{w \in W} \rangle$ *pour* $\mathcal{L}_{\Box\!\!\rightarrow}$ *est un n-uplet constitué*

1. *d'un ensemble W de mondes possibles*
2. *d'une valuation atomique $I : At \rightarrow \wp(W)$*
3. *d'un sytème de sphères $(\mathcal{S}_w)_{w \in W}$.*

Définition 5 (Relation de satisfaction)
Soit $M = \langle W, I, \mathcal{S} \rangle$ un modèle de sphères ; pour tout $w \in W$, on définit ainsi la relation de satisfaction :

(i) $M, w \models p$ *ssi* $w \in I(p)$
(ii) $M, w \models \neg\phi$ *ssi* $M, w \not\models \phi$
(iii) $M, w \models (\phi \wedge \psi)$ *ssi* $M, w \models \phi$ *et* $M, w \models \psi$
$M, w \models (\phi \vee \psi)$ *ssi* $M, w \models \phi$ *ou* $M, w \models \psi$
$M, w \models (\phi \rightarrow \psi)$ *ssi si* $M, w \models \phi$ *alors* $M, w \models \psi$
(iv) $M, w \models (\phi \Box\!\!\rightarrow \psi)$ *ssi s'il existe une sphère* $S \in \mathcal{S}_w$ *telle que* $[[\phi]] \cap S \neq \emptyset$, *alors il existe une sphère* $T \in \mathcal{S}_w$ *telle que* $[[\phi]] \cap T \neq \emptyset$ *et pour tout* $v \in T$, $M, v \models (\phi \rightarrow \psi)$.

Voir FIGURE 3 de Lewis (1973a).

Remarque : Lewis appelle les sphères où ϕ est vraie dans certains mondes les sphères ϕ-*permitting*. La clause (iv) dit que s'il existe une sphère ϕ-*permitting* dans $\mathcal{S}_w$, alors $(\phi \rightarrow \psi)$ est vraie dans tous les mondes d'une sphère ϕ-*permitting*.

Définition 6
Une formule ϕ est valide dans les modèles de sphères ssi elle est vraie en tout monde de tout modèle de sphères ; on le note

$$\boxed{\models_{sph} \phi}$$

5 Syntaxe

Lewis construit un système axiomatique pour les modèles de similitude (et, par conséquent, pour les modèles de sphères) : le système **VC**.

$(\phi \preceq \psi) =_{df} \neg(\psi \prec \phi)$ $\Diamond\phi =_{df} (\phi \prec \bot)$ $\Box\phi =_{df} \neg(\Box\neg\phi)$ $(\phi \Box\!\!\rightarrow \psi) =_{df} ((\Diamond\phi \rightarrow ((\phi \wedge \psi) \prec (\phi \wedge \psi)))$
(PROP) Instances de tautologies (MP) De ϕ et $(\phi \rightarrow \psi)$ inférer ψ
$((\phi \preceq \psi \preceq \chi) \rightarrow (\phi \preceq \chi))$ $(\phi \preceq \psi) \vee (\psi \preceq \phi))$ $((\phi \preceq (\phi \vee \psi) \vee (\psi \preceq (\phi \vee \psi)))$
(C) $((\phi \wedge \neg\psi) \rightarrow (\phi \prec \psi))$
De $(\phi \rightarrow \psi)$ inférer $(\psi \preceq \phi)$

• On peut passer du système **VC** à un système équivalent au système **C2** de Stalnaker en ajoutant à **VC** l'axiome du tiers-exclu conditionnel :

(CEM) $((\phi \Box\!\!\rightarrow \psi) \vee (\phi \Box\!\!\rightarrow \neg\psi))$

Théorème 1 (Théorème de complétude)

$$\boxed{\vdash_{VC} \phi \textit{ ssi } \models_{sim} \phi \textit{ ssi } \models_{sph} \phi}$$