Raisonnements dans les systèmes logiques formels et leur représentativité des systèmes cognitifs naturels

Un essai d'étude des trois modes de dérivation des connaissances : déduction, induction, abduction

Violaine Prince
Professeur à l'Université Paris VIII

1998

I-Déduction

La déduction est un mode de dérivation de propositions logiques utilisé depuis fort longtemps. C'est son mécanisme qui sert de base aux syllogismes prisés des Grecs. Il a été formalisé dans sa structure actuelle par Descartes. Nous en donnons brièvement les fondements.

I.1 Conditions initiales

La déduction suppose que nous soyions dans l'univers de la logique des propositions de laquelle on connaît les propriétés suivantes.
Chaque proposition peut recevoir, de manière certaine, une valeur d'interprétation, puisée dans l'ensemble {vrai, faux}.
On désignera les propositions par des lettres A, B, C, etc.
On a à disposition les connecteurs logiques qui sont :

- $\wedge$ qui symbolise le «et» logique et qui se comporte comme la multiplication en arithmétique
- $\vee$ qui est le «ou» logique et qui se comporte comme l'addition en arithmétique
- $\supset$ ou $\Rightarrow$ qui est l'implication logique, et qui ne représente pas forcément l'implication causale, ni l'antécédence chronologique (la confusion est malheureusement trop systématique dès lors que l'on essaie de représenter en logique des propositions des éléments émanant d'une réalité)
- $\neg$ qui est la négation logique, et qui inverse l'interprétation d'une proposition (si cette proposition est vraie, sa forme niée est fausse et réciproquement).

Les opérateurs logiques sont des relations (au sens ensembliste du terme) entre propositions mais des relations évaluées avec une certaine sémantique. Cette évaluation est donnée par les tables de vérité. Ainsi, la formule $A \wedge B$ est une proposition dont l'interprétation dépend des valeurs prises par A et par B. La table de vérité correspondante dénote la sémantique de l'opérateur «et» et donne, par exemple :

A	B	$A \wedge B$
vrai	vrai	vrai
vrai	faux	faux
faux	vrai	faux
faux	faux	faux

I.2 Le mécanisme de déduction

Le mécanisme repose sur la propriété de transitivité de l'opérateur «implication». Il s'écrit comme suit :
$(A \supset B \wedge B \supset C) \supset (A \supset C)$
ce qui signifie : «A implique B et B implique C implique «A implique C».

Le problème est que cette écriture est très ambiguë. En effet, la proposition résultant de cette composition de propositions est nécessairement toujours vraie quelle que soit la valeur attribuable aux propositions A, B et C, et surtout, ce qui est plus grave, si la partie gauche, $A \supset B \wedge B \supset C$ que l'on appelle la *prémisse*, est fausse, alors que la partie droite $A \supset C$, appelée *conclusion*, est vraie. On peut d'ailleurs faire la démonstration du mécanisme de déduction, puisque la proposition résultante, que l'on nommera X1, est interprétée grâce à la sémantique du connecteur $\supset$.

Démonstration

Cette sémantique est fournie par la table de vérité suivante, qui sert de démonstration :

A	B	$A \supset B$	C	$B \supset C$	$A \supset B \wedge B \supset C$	$A \supset C$	X1
vrai	vrai	vrai	vrai	vrai	vrai	vrai	vrai
vrai	faux	faux	vrai	vrai	faux	vrai	vrai
vrai	faux	faux	faux	vrai	faux	faux	vrai
vrai	vrai	vrai	faux	faux	faux	faux	vrai
faux	vrai	vrai	vrai	vrai	vrai	vrai	vrai
faux	vrai	vrai	faux	faux	faux	vrai	vrai
faux	faux	vrai	vrai	vrai	vrai	vrai	vrai
faux	faux	vrai	faux	vrai	vrai	vrai	vrai

Table 1. Table de vérité démontrant le théorème de déduction.

La proposition X1 prend toujours la valeur «vrai». C'est donc un théorème de tout système logique fondé sur le calcul propositionnel qui comprend l'implication.

Commentaire :

Le problème épistémologique que l'on détecte lorsque l'on pose le mécanisme de déduction *in abstracto* revient à rejeter l'idée intuitive selon laquelle la déduction puisse se faire alors que la prémisse est fausse. Or la sémantique de $\supset$ est telle qu'une prémisse fausse peut impliquer n'importe quoi au sens logique. En effet $A \supset B$ se comporte comme $\neg A \vee B$ qui est toujours vrai sauf dans le cas ou A est vrai et B est faux, ce qui signifie que le vrai n'implique jamais le faux, alors que le faux peut impliquer aussi bien le vrai que le faux. C'est comme cela que l'on peut déduire des propositions vraies à partir de prémisses fausses.

Le théorème de déduction est en effet applicable même si $A \supset B$ est faux et/ou si $B \supset C$ est faux ! On voit donc tout de suite le danger de l'usage du théorème en dehors de toute condition de validité dès lors que la logique sert à dénoter la réalité. Ce que l'on peut appeler une faute de raisonnement, au sens commun de ce terme, n'est pas une faute dans un système logique où seul le théorème de déduction existe.

C'est pourquoi, tout système modélisateur (c'est-à-dire qui veut représenter une réalité) doit non pas appliquer le théorème de déduction *in abstracto*, mais vérifier les seuls cas acceptables et qui sont ceux où $A \supset B \wedge B \supset C$, (par défaut, la proposition écrite représente la valeur «vraie» attribuée à cette proposition) c'est-à-dire lorsque $A \supset B$ et $B \supset C$ sont simultanément vrais. Cela sélectionne donc les lignes suivantes de la table de vérité précédente :

A	B	$A \supset B$	C	$B \supset C$	$A \supset B \wedge B \supset C$	$A \supset C$	X1
vrai	vrai	vrai	vrai	vrai	vrai	vrai	vrai
faux	vrai	vrai	vrai	vrai	vrai	vrai	vrai
faux	faux	vrai	vrai	vrai	vrai	vrai	vrai
faux	faux	vrai	faux	vrai	vrai	vrai	vrai

Table 2. Restriction de la déduction aux seules prémisses vraies.

I.3 Le raisonnement déductif

Pour faire un véritable raisonnement, c'est-à-dire acceptable sur le plan de la démonstration, on remplacera le théorème de déduction tout seul par le système suivant qui aura pour nom «raisonnement déductif» :

1. $A \supset B$
2. $B \supset C$
3. $(A \supset B \wedge B \supset C) \supset (A \supset C)$

Figure 1. Représentation de l'axiomatique du raisonnement déductif.

qui signifie que le théorème de déduction est applicable dans un système dans lequel les propositions 1 et 2 sont vraies (et non démontrées). C'est une contrainte *sine qua non*, parce que sinon, la tentation de déduire $A \supset C$ de prémisses fausses est très grande, et ce sans vérification préalable de la valeur prise par $A \supset C$. C'est ce qui va poser les premières tentatives d'axiomes inductifs logiques que nous verrons dans la deuxième section.

I.4 La représentativité du système déductif

La représentativité d'un modèle dénote jusqu'à quel point ce modèle peut se «substituer» à la réalité de référence sans profondément défigurer celle-ci.
Le problème du système déductif est que, bien qu'il soit fort employé, sa représentativité est faible. L'univers de la déduction pure est celui de la logique des propositions. Cela signifie donc que l'on devrait savoir, de source sûre, attribuer une valeur de vérité à des objets correctement formulés. La réalité de beaucoup de disciplines scientifiques est malheureusement autre. Si certains domaines peuvent supporter sans dommage une modélisation en terme de logique des propositions, parce que l'interprétation «vrai, faux» y est possible et suffisamment constante, la plupart des connaissances humaines, dès lors qu'elles font intervenir des systèmes ouverts, ne possèdent pas les bonnes conditions pour être représentés par cette logique.
Outre la question de l'attribution de la valeur de vérité se pose celle de la sémantique des connecteurs, l'élément le plus représentatif du problème étant justement le connecteur d'implication.

I.4.1Implication logique et implication causale

Comme on a pu le voir, l'implication logique n'est pas obligatoirement le modèle idoine pour l'implication causale. «A est cause de B» ne se représente pas forcément par $A \supset B$. Etudions pour cela la notion de causalité. La proposition «A est cause de B» signifie que si B est vraie (B est apparu) alors A peut être vrai aussi. Si nous regardons dans la table 1 les valeurs de $A \supset B$ quand B est vrai, on s'aperçoit que l'on obtient la table suivante :

A	B	$A \supset B$
vrai	vrai	vrai
vrai	vrai	vrai
faux	vrai	vrai
faux	vrai	vrai

Table 3. valeurs de $A \supset B$ et A quand B est vrai.

Comme on le sait, A peut parfaitement être faux, l'implication logique est vraie quand même. Ce que $A \supset B$ représente c'est que A peut être une des causes possibles de B, mais pour que A puisse être raisonnablement envisagé comme cause quand B est vrai, seule la formule

$A \wedge A \supset B$

qui signifie que B est vrai et A est vrai simultanément le permet. On peut effectivement le vérifier à l'aide de la table 4 :

A	B	$A \supset B$	$A \wedge A \supset B$
vrai	vrai	vrai	vrai
vrai	faux	faux	faux
faux	vrai	vrai	faux
faux	faux	vrai	faux

Table 4. Table de vérité de l'implication modulée par des prémisses vraies.

Donc effectivement, le seul cas où $A \wedge A \supset B$ est vrai c'est lorsque A et B sont simultanément vrais. On remarquera qu'il faut donc, dans un système déductif, appliquer de façon récursive les contraintes de vérité, ce qui nous amène aux conditions idéales de la déduction pure qui sont tout simplement :

$A \wedge B \wedge C$

à partir de là, la déduction s'applique, même quand A n'est pas une cause de B, et B n'est pas une cause de C. Ainsi trois propositions simultanément vraies peuvent être déduites l'une de l'autre même quand elles n'ont pas de relation causale entre elles !

Une corruption de cette constatation peut être ainsi illustrée par l'exemple suivant .

A=Tous les hommes de plus de 1m85 sont grands.

B=Le Danemark est le pays de la liberté intégrale de la presse.

C=Le taux de suicides en Suède est parmi les plus élevés du monde.

Si on applique les seules règles de la déduction pure, et que l'on confonde implication logique et implication causale, on peut parfaitement montrer que :

Tous les hommes de plus de 1m85 sont grands implique que le Danemark est le pays de la liberté intégrale de la presse
Le Danemark est le pays de la liberté intégrale de la presse implique que le taux de suicides en Suède est parmi les plus élevés du monde.
On en déduit donc que :
Tous les hommes de plus de 1m85 sont grands implique que le taux de suicides en Suède est parmi les plus élevés du monde.

Bien sûr cela paraît absurde énoncé comme cela, parce que nous savons parfaitement que les choses ne sont pas liées de façon causale, mais si on avait été plus «vicieux» en présentant les propositions suivantes :
La Suède et le Danemark ont un climat très froid.
Le Danemark est le pays de la liberté intégrale de la presse.
Le taux de suicides en Suède est parmi les plus élevés du monde.
On serait très tenté de dire que la rigueur du climat entraîne le laxisme d'opinion, et la dépression suicidaire. C'est peut-être incidemment vrai, mais on ne peut pas le démontrer, alors que l'implication logique, elle, est parfaitement démontrable en logique des propositions à partir des valeurs de la table de vérité.
En conclusion, nous avons pu constater que l'implication logique n'entraîne pas forcément une implication causale, et inversement, une implication causale, si elle doit être représentée par une implication logqiue, impose nécessairement la validation de la prémisse d'une part, mais est-ce que cette validation est suffisante ? A priori non, sauf dans le cas du raisonnement déductif (voir figure 1).

Les gens qui font de la modélisation à l'aide de la logique des propositions doivent être attentifs à ce genre d'anomalie lorsqu'ils vont manuellement représenter des implications causales naturelles (c'est-à-dire d'un domaine de connaissances de l'être humain sur le monde) par des implications logiques. C'est effectivement le cas des personnes qui représentent des connaissances dans des systèmes d'Intelligence Artificielle avec le formalisme des règles de production qui relève d'une écriture sous forme d'implication logique (les détails techniques de l'écriture seront introduits dans la suite de notre propos). Le problème n'est donc pas au moment où les auteurs écrivent leurs système. C'est au moment où le système va augmenter l'ensemble des connaissances. En général, ce qui est autorisé, c'est uniquement le raisonnement déductif tel que nous l'avons donné dans la figure 1. Si les prémisses sont elles-mêmes des implications et qu'elles sont vraies, si la prémisse de la seconde implication correspond à une conclusion de la première, alors on applique le théorème de déduction, et on produit une nouvelle implication par transitivité. Cette méthode-là est fiable, et marche à cent pour cent. En revanche, il faut totalement proscrire $A \supset B$ en tant que résultat, lorsque A et B sont vrais et que rien ne permet d'appliquer la transitivité de la relation «implique». Là on tombe dans des déductions abusives du genre de celles que nous avons données en exemple.
Ces limites imposées à un système automatique fondé sur le calcul des propositions doivent être fournies en tant qu'axiomes de circonscription de ce système. Sinon, il risque de générer des connaissances «anormales» en particulier lorsque le modèle retenu pour représenter l'implication causale est celui de l'implication logique.

I.4.2 L' implication logique, modèle ambigu

Outre le fait que l'implication logique représente assez mal la sémantique de la causalité du sens commun, malheureusement, elle reste le seul moyen de représenter les connaissances de type «si…alors» qui elles-mêmes sont de statut épistémologique divers. C'est Claude Vogel (1988) qui a le plus clairement énoncé ces différences, tandis qu'il proposait une méthode pour modéliser les connaissances des experts (humains) en vue de les programmer dans des systèmes à base de calcul logique.
Les différentes situations épistémiques qui peuvent être représentées par les connaissances formulées par «si…alors» sont les suivantes :

- la causalité

exemple :

si les bougies sont encrassées alors le moteur aura du mal à démarrer

- les conditions d'une action

exemple :

si le voyant d'huile est allumé alors il faut rajouter immédiatement de l'huile à moteur

- la succession chronologique

exemple :

si le train n'est plus annoncé sur le panneau d'affichage des départs, alors il est déjà parti depuis un moment.

Le problème est qu'il n'y a pas non plus unicité de la formulation en langue de ces différentes situations épistémiques : la présence du «si…alors» est facultative, et elle peut être remplacée par «quand» ou «dès que», pas seulement pour le seul cas de succession chronologique. On aurait pu parfaitement dire :

quand les bougies sont encrassées le moteur a du mal à démarrer

avec la concordance de temps qui s'impose.

I.5 La représentativité de la logique des propositions et de l'implication logique

La transformation de ces énoncés en langue en propositions logiques va poser plusieurs problèmes que les nuances de la langue escamotent, et qui apparaissent tout crûment en logique.Exemple :

si les bougies sont encrassées alors le moteur aura du mal à démarrer

est transformé en :

les bougies sont encrassées $\supset \neg$*(le moteur démarre)*

on remarque que la nuance «a du mal» disparaît, et que le conditionnel aussi ne vient plus atténuer la conclusion. Pour restituer ce genre de nuances, les concepteurs de systèmes à base de connaissance ont attribué soit une probabilité à la règle (l'ensemble de l'énoncé propositionnel), soit une probabilité à la conclusion (si la prémisse est vraie). Dans le premier cas, on est dans une représentation à la MYCIN, dans le deuxième, on peut transformer cet énoncé en logique floue avec une fonction de croyance associée.

Cependant, les erreurs d'interprétation ne sont pas seulement là : des bougies peuvent être encrassées, mais pas suffisamment pour être la cause du non démarrage du moteur. Le moteur peut soit bien démarrer en dépit de ses bougies, et du coup, l'implication est invalide (la conclusion est fausse), soit au contraire ne pas démarrer (la conclusion, donc l'implication sont valides) mais pas du tout à cause de ses bougies. Examinons les deux cas.

I.5.1 La falsification d'une connaissance

Le premier cas va invalider l'implication par invalidation de la conclusion, c'est-à-dire rendre fausse la règle :

les bougies sont encrassées $\supset \neg$*(le moteur démarre)*

Si c'est une connaissance incluse dans la base des connaissances du système programmé, alors il va y avoir un conflit entre le fait constaté, qui peut être «le moteur démarre quand même», et les connaissances de la base.
Le deuxième cas valide l'implication, mais on peut s'apercevoir que le fait «le moteur ne démarre pas» ne provoque éventuellement des conflits qu'à la suite du déroulement d'un plan d'actions, c'est-à-dire par enchaînement d'actions que nous étudierons dans le prochain paragraphe.
Nous nous limitons ici à la constatation d'un démarrage du moteur (falsification de la connaissance), mais un démarrage cahotant, qui correspond bien à la nuance de l'énoncé en langue («avoir du mal») mais qui n'est plus restitué dans la forme logique.
Un être humain saurait très bien gérer ce conflit entre une règle de ses connaissances, et le fait constaté : étant réflexif, c'est-à-dire capable d'évaluation, il va temporairement mettre de côté cette règle en la considérant comme localement non pertinente. Un système à base de connaissances fondé sur la logique des propositions (les prédicats du premier ordre en sont une écriture plus souple, mais pas fondamentalement différente) n'a pas les moyens de mettre en oeuvre ce genre de mécanisme évaluatif. L'invalidation d'une règle doit entraîner sa disparition du système, parce que ce dernier ne peut effectuer de la déduction sans aberration que sur des connaissances vraies, ainsi que nous l'avons montré en I-2. Le fait constaté, étant sans nuance dans sa traduction logique, tout comme la conclusion, entraînera *de facto* soit la révision de la base de connaissances par retrait de la règle, soit le rejet du fait. Les deux actions sont évidemment grandement dommageables, parce que la règle peut être valide dans un autre contexte, et le fait n'est pas faux, mais simplement mal traduit.

I.5.2 La falsification d'un plan ou d'un résultat

Le deuxième cas est plus complexe, parce qu'il va entraîner des raisonnements déductifs par chaînage arrière, mais une recherche de raisonnement par chaînage avant.

On constate que le moteur ne démarre pas. On applique la règle correspondant aux connaissances, et on cherche donc à savoir si les bougies sont en cause : en effet, la prémisse «les bougies sont encrassées» doit être alors étudiée dans ses valeurs de vérité. Le **chaînage arrière** consiste à regarder de quelles règles cette prémisse est elle-même la conclusion pour remonter ainsi à une situation de diagnostic. Dans le cas des applications de type diagnostic pur, un exemple de chaînage peut être :

les bougies sont encrassées $\supset \neg$*(le moteur démarre)*
les bougies sont vieilles $\supset$ *les bougies sont encrassées*

on en conclut que la vieillesse des bougies peut être une cause de leur encrassement. Le fonctionnement en tant que tel est satisfaisant, encore que des bougies puissent être encrassées sans être vieilles et donc le résultat peut très bien être falsifié par la réalité. Les conclusions d'un tel système de diagnostic sont vraies par rapport à sa base de connaissances mais peuvent parfaitement être fausses en tant que représentation de la réalité. Quelle en est la cause ? Souvent l'incomplétude des connaissances de la base. Pour ne pas être en reste par rapport à la réalité, il aurait fallu représenter dans la base tous les cas où des bougies pourraient être encrassées. Dès lors, le diagnostic du système ne consiste pas en un résultat atomique, mais un résultat ensembliste, dans lequel on se donne toutes les chances pour qu'au moins une des propositions du système s'apparie avec la réalité.
Si on a affaire à un monde de connaissances circonscrit ou fermé, il est possible de créer une base complète. Si on a affaire à un monde ouvert, l'incomplétude est intrinsèque. C'est **le** problème de fond de la représentation des connaissances par des systèmes logiques.

Dans le cas des applications où l'ensemble des actions à effectuer doit être préconisé, le **chaînage avant**, qui consiste à regarder quelles sont les règles ayant les mêmes prémisses et à les essayer, est plus performant. Prenons par exemple ces propositions appartenant à la base de connaissances :

les bougies sont encrassées $\supset \neg$*(le moteur démarre)*
les bougies sont encrassées $\supset$ *changer les bougies*

Le problème apparaît lorsque l'application est mixte, c'est-à-dire diagnostic-action, ou encore, lorsque l'on est obligé de mélanger ces deux situations épistémiques qui sont décrites par le même formalisme. En pure logique, une prémisse vraie implique une conclusion vraie. Comme ici la conclusion est une action, alors le plan consistant à réaliser les actions découlant d'une prémisse vraie est proposé comme solution. Cela fonctionne sur le plan formel lorsqu'il y a peu de règles avec des prémisses communes (comme c'est le cas pour cet exemple). Cela devient plus problématique lorsque plusieurs règles décrivant des aspects différents de la réalité, ont des prémisses (ou conditions) en commun. De toutes façons, plusieurs cas de figure peuvent apparaître :

- le plan exécuté ne résout pas le problème : ici, le fait de changer les bougies ne permet pas de faire redémarrer le moteur, il faut donc repartir de cette constatation pour trouver d'autres prémisses de règles qui iraient vers d'autres plans
- le plan n'est pas unique car il propose des diagnostics (conclusions) et des actions multiples selon la richesse des règles intégrées dans la base : se pose alors le problème de retour arrière (ou back-track) pour choisir dans l'arborescence des plans
- le plan ne se hiérarchise pas en actions structurées : les actions seront menées un peu de façon arbitraire, et tous le bénéfice d'une organisation rationnelle des

connaissances est quelque peu terni par un plan qui apparaît comme un ensemble de recettes de cuisine à essayer dans un ordre quelconque.

Dans tous ces cas, la confrontation avec la réalité falsifie soit le plan soit le résultat de celui-ci, et cela, parce que la base de connaissances est peu représentative de la réalité qu'elle décrit. La question pour nous n'est pas seulement celle du nombre de connaissances représentées dans la base, mais des limites intrinsèques de la logique des propositions et de l'implication quand elles traduisent des connaissances naturelles.

1.5.3 La représentativité cognitive du Modus Ponens

Le Modus Ponens est un des axiomes fondamentaux des systèmes déductifs qui permet justement de produire des connaissances. Il va s'écrire de la façon suivante :

$$\frac{A \wedge A \supset B}{B}$$

ce qui signifie que si A est vrai et si $A \supset B$ est vrai alors B est vrai. On remarquera qu'il correspond aux conclusions que nous avons faites dans le paragraphe I-4.1 (et la table 4). C'est effectivement lui que l'on applique en chaînage avant, c'est-à-dire en regardant si une prémisse est vraie et en faisant l'hypothèse que la règle de la base est vraie. C'est le Modus Ponens qui bloque lorsque l'on constate $\neg B$.
Mendelsohn (1964) a montré que le système formel composé des trois axiomes suivants :

$$A \supset (B \supset A)$$
$$(A \supset (B \supset C)) \supset ((A \supset B) \supset (A \supset C))$$
$$(\neg B \supset \neg A) \supset ((\neg B \supset A) \supset B)$$

et du Modus Ponens, est un **système de déduction complet**[1], et que l'on peut par ailleurs l'utiliser pour démontrer les trois règles d'inférence suivantes :

1) $\frac{A}{B \supset A}$ qui signifie que si A est vrai, alors pour n'importe quelle proposition B, $B \supset A$ est vrai. Ce qui explique que l'on puisse trouver des causes qui n'en sont pas (non pertinence), et même des causes fausses puisque $B \supset A$ peut-être vrai même si B est faux.

2) $\frac{(A \supset B) \wedge (B \supset C)}{A \supset C}$ qui est la propriété de transitivité que nous avons postulée comme fondamentale. A notre avis, le deuxième axiome et le Modus Ponens sont une réécriture de cette propriété.

[1] On dit qu'un système déductif est complet si toute proposition valide est déductible du système. On dit qu'une proposition B est déductible d'un système s'il existe une séquence finie de propositions telle que B est la dernière séquence, et toutes les autres séquences sont soit des axiomes du système, soit des proprositions déduites du système.

3) $\dfrac{A \supset (B \supset C)}{B \supset (A \supset C)}$ qui montre que si A est la prémisse d'une règle qui a pour conclusion une autre règle, et que l'on observe ou suppose que cette « méta-règle » est vraie, alors on peut montrer que l'on peut échanger les prémisses, et que la chaîne des « causes » est commutative.

En d'autres termes, les systèmes complets de déduction, même aussi simples que celui de Mendelsohn qui ne comprend que deux connecteurs, la négation et l'implication, sont des systèmes qui sont dangereux pour la représentation des connaissances naturelles. Ni la négation logique ni l'implication logique ne représentent les connaissances négatives, respectivement les connaissances causales, ou ordonnées. Nous pourrons discuter plus en avant des propriétés de la conjonction (et) et de la disjonction (ou) logiques, qui ne recouvrent que partiellement, voire à contresens la conjonction et la disjonction « naturelles » (telles qu'elles sont exprimées par les langues).
Et de plus, la représentativité cognitive du Modus Ponens est faible puisqu'elle est fondée sur les propriétés fondamentales de l'implication logique, chose que nous avons essayé de montrer dans les paragraphes I-4 et I-5.

II-Induction

L'induction est le mode de raisonnement qui permet de passer de l'élément vers l'ensemble, et ce, grâce à la répétitivité constatée de certaines propriétés des éléments. Par exemple, la constitution d'une classe d'équivalence de X, $\overline{X}$, qui est composée de tous les éléments x_i tels que $(x_i \equiv X)_R$ dans un système d'interprétation quelconque (il existe une relation R réflexive, symétrique et transitive permettant cette équivalence), est le résultat d'un raisonnement inductif.
La représentation d'un ensemble par sa définition intensionnelle à partir d'une définition extensionnelle est le résultat de l'application de l'induction.
Ainsi l'équivalence entre $\{1,2,3,4,5,6,7,8,9,10\}$ et $\{i \in [1,10] \subset N\}$ est postulable parce que les propriétés (définition intensionnelle) décrivent parfaitement les éléments (définition extensionnelle). L'induction est tellement naturelle qu'on n'y a pas prêté attention.

Dans un système logique, il ne faut pas confondre induction et inférence, car certaines inférences sont inductives et d'autres sont déductives, et enfin d'autres sont abductives.
Une inférence déduite est celle représentée par l'application de la transitivité de l'implication logique, ou par un système axiomatique muni du Modus Ponens, tels que nous les avons vus dans le chapitre précédent.
Une inférence relevant de l'induction est le résultat d'une classification réussie, c'est pourquoi, l'induction est surtout utilisée dans les logiques quantifiées, c'est-à-dire à partir du moment où l'on peut utiliser les quantificateurs universel et existentiel.

II-1 Induction logique : les axiomes, la validité, les limites

L'induction logique va demeurer dans les limites des systèmes déductifs complets. Gödel a montré que la logique des prédicats du premier ordre, comme celle des propositions, possédait des systèmes déductifs complets (Premier théorème). Pour présenter l'induction, nous allons fonctionner avec un système intermédiaire, celui des propositions quantifiées. On sait de toutes façons trouver une correspondance entre une formule quantifiée et une formule prédicative du premier ordre. Cependant, avant de travailler sur l'induction, nous allons décrire un système pré-inductif qui est le système hypothético-déductif.

II.1.1 Révision de la déduction en ajoutant la structure hypothétique : le système dit « hypothético-déductif »

Le système hypothético-déductif permet la révision des connaissances. En effet, il va considérer les connaissances sur le monde comme ré-interprétables, moyennant un ensemble donné de connaissances acquises. Cela signifie qu'on ne travaille plus sur un ensemble fixe de connaissances, mais sur un ensemble évolutif. C'est lui qui a servi intuitivement dans notre évaluation de la falsification des résultats et des plans dans les paragraphes I-4 et I-5. Ici, nous allons essayer de le formaliser.

A partir de maintenant, on va supposer que chaque proposition, qu'elle soit atomique comme (« le chat est mort », représenté par la simple lettre A) ou formule comme ($B \supset A$) est une **hypothèse**, et que l'on possède **un ensemble fini d'hypothèses** appelé Γ. On part donc d'un système déjà initialisé non vide comprenant des propositions vraies tant qu'on n'a pas montré qu'elles étaient fausses. Cela est déjà différent d'un système où les propositions sont vraies ou fausses sans «pari» hypothétique, ce qui était le cas du système déductif pur.

Pour faire évoluer le système de connaissances, on possède un moyen d'action sur ces propositions qui consiste à en ajouter ou à en enlever.
On va supposer que l'adjonction d'une hypothèse A à Γ s'écrit Γ, A et que la formule qui consiste à écrire :
{toutes les hypothèses de Γ ∧ A} ⊃B va s'écrire Γ, A⇒B. Cette écriture est adoptée de (Manna , 1974). C'est une **implication ensembliste** et généralisée. Elle permet de comparer l'état des connaissances à chaque étape avec l'état du système de l'étape antérieure.
On va aussi représenter le mode d'acquisition des connaissances par adjonction ou soustraction «déduites» par le symbole de barre de fraction. Généralisé à l'implication ensembliste, ce symbole signifie qu'en prémisse (numérateur) il y a une implication ensembliste et en conclusion (dénominateur) il y a une autre implication ensembliste, déductible de la précédente, dont l'objectif est de faire évoluer l'ensemble des connaissances.

On décrit alors les principaux axiomes du système hypothético-déductif dans la figure 2.L'ensemble des axiomes est repris à Manna.

1) $\Gamma, A \Rightarrow A$ si A est une hypothèse du système, alors elle est déductible de lui. (**axiome de l'hypothétisation**).

2) $\dfrac{\Gamma \Rightarrow B}{\Gamma, A \Rightarrow B}$ si B est déductible du système, alors B est déductible de tout surensemble d'hypothèses du système. (**adjonction d'une hypothèse**)

3) $\dfrac{\Gamma, A \Rightarrow B \;\; et \;\; \Gamma, \neg A \Rightarrow B}{\Gamma \Rightarrow B}$, si A est une hypothèse du système, et si B est déductible du système quelle que soit la validité de A, alors A doit être éliminée du système. (**rétraction d'une hypothèse**)

4) $\dfrac{\Gamma, A \Rightarrow B}{\Gamma \Rightarrow A \supset B}$ Si B est déductible d'un ensemble d'hypothèses et d'une hypothèse A alors la règle A⊃B est déductible du système. (**adjonction d'une implication**)

5) $\dfrac{\Gamma \Rightarrow A \;\; et \;\; \Gamma \Rightarrow A \supset B}{\Gamma \Rightarrow B}$. Si A est déductible du système, et si A est prémisse d'une règle déductible du système, alors la conclusion de cette règle est directement déductible du système (**rétraction d'une implication, modus ponens**).

Figure 2. Axiomes du système hypothético-déductif.

On remarquera que la rétraction d'hypothèse (axiome 3) n'élimine pas les hypothèses non valides, mais uniquement les hypothèses non pertinentes. C'est donc un problème d'optimisation du nombre d'hypothèses du système, hypothèses qui peuvent avoir été ajoutées de façon non contrôlée par l'axiome 2.
La règle 4 est typique des systèmes logiques, c'est une réécriture de $((X \wedge A) \supset B) \supset (X \supset (A \supset B))$ pour tous les X de Γ. Le problème est qu'elle risque de générer des règles qui n'ont rien à voir avec l'implication causale.
Enfin on constatera que la règle 5 ne retire une implication que si elle est superflue, et non pas si elle est remise en cause. En effet, le Modus Ponens du système hypothético-déductif retirera les règles redondantes. Une gestion de la cohérence par ces axiomes est en réalité régie par le souci d'économie, mais pas par le souci de représentativité.
La figure 3 permet, grâce à la négation, de gérer les incohérences logiques (contradictions).

6) $\dfrac{\Gamma, A \Rightarrow B \;\; et \;\; \Gamma, A \Rightarrow \neg B}{\Gamma \Rightarrow \neg A}$ si B est déductible du système muni de l'hypothèse A, et si B est aussi falsifiable par le système muni de A, alors c'est A qui est falsifiée. (règle dite de **reductio ad absurdum**)

7) $\dfrac{\Gamma \Rightarrow A \;\; et \;\; \Gamma \Rightarrow \neg A}{\Gamma \Rightarrow F}$ si A et non A sont déductibles d'un même système alors ce système est faux (**élimination de la négation**).

Figure 3. Règles de gestion des incohérences, ou règles de négation.

On remarquera que l'axiome 6 suppose que Γ reste cohérent, chose vérifée par l'axiome 7 qui montre que si on n'arrive pas à mettre le doigt sur l'hypothèse qui introduit l'incohérence (celle de l'axiome 6) alors c'est la totalité du système hypothétique qui est est remis en cause.
Cela signifie que le système hypothético-déductif sait partiellement gérer les incohérences, mais ne sait pas gérer l'adéquation à la réalité. Il ne peut donc proposer que des pistes de diagnostic cognitif (savoir/croire ou non), sans pour autant certifier ce diagnostic. Il serait encore plus hasardeux de le transformer en système « actif ».

II.1.2 Le Modus Tollens ou la réfutation des causes

Le Modus Tollens est le raisonnement qui mettra en oeuvre la gestion des implications par un système de connaissances révisables, et qui sera sollicité pour débloquer le Modus Ponens. En effet, ce raisonnement stipule que si un système possède une implication et que, en raison d'un fait ou d'un autre raisonnement, la conclusion de cette implication est falsifiée, alors c'est la prémisse de la règle qui est invalidée.
La formule du Modus Tollens est la suivante :

$$\frac{\Gamma \Rightarrow A \supset B \;\; et \;\; \Gamma \Rightarrow \neg B}{\Gamma \Rightarrow \neg A}$$

reprenons l'exemple proposé dans le premier chapitre, et considérons l'ensemble de «connaissances» hypothétique suivant :

les bougies sont encrassées ⊃ ¬(le moteur démarre)

les bougies sont vieilles $\supset$ *les bougies sont encrassées*
les bougies sont encrassées $\supset$ *Il faut changer les bougies*

Si le moteur démarre, alors le Modus Tollens peut déduire que les bougies ne sont pas encrassées, ce qui peut être faux dans la réalité, mais au fond sans grand dommage puisque cela n'empêche pas le moteur de démarrer, et le raisonnement se solde par une non action, même si l'ensemble cognitif n'adhère pas obligatoirement à la réalité. Mais dans un ensemble du type :

les bougies sont neuves $\supset$ *le moteur démarre*
$\neg$*(lesbougies sont neuves)* $\supset$ *les bougies sont encrassées*
les bougies sont neuves $\supset$ $\neg$*(Il faut changer les bougies)*

qui est une façon a priori équivalente du point de vue logique à la précédente, le démarrage du moteur va mettre le système dans un état cognitif non forcément adéquat mais sans dommage. En revanche, le non-démarrage du moteur qui, dans le premier cas, va conduire à changer les bougies (sans pouvoir réellement évaluer si l'action est efficace ou non) va invalider la prémisse «les bougies sont neuves» mais va complètement bloquer l'action. La troisième règle ne peut pas s'appliquer puisque le Modus Ponens impose que la prémisse soit vraie. Donc le système ne produira pas d'action parce que sa représentation des connaissances n'est pas adéquate avec un raisonnement par Modus Tollens.

En résumé, concernant les systèmes hypothético-déductifs qui, *a priori*, sont plus adaptés à la cognition humaine que les systèmes déductifs purs, les problèmes résiduels sont :

- une connaissance retirée l'est exclusivement pour des raisons de redondance ou de réduction de l'absurdité.
- La négation partielle n'est pas représentée.
- La négation hypothétique non plus.
- L'adéquation au réel ne fait pas partie du système : une implication logique qui n'est pas causale n'échappe pas au Modus Ponens.
- La réfutation des causes se fait cause par cause : une conclusion fausse entraîne la falsification des causes par Modus Tollens, et la notion de «cause hypothétique» n'existe pas.

II.1.3 Le système hypothético-déductif quantifié : introduction des axiomes inductifs avec la quantification universelle

C'est en réalité à partir du système hypothético-déductif quantifié que l'on est en mesure d'introduire les mécanismes inductifs en logique. En effet, la quantification va traiter de points qui n'ont pas été directement révisables par le système hypothético-déductif, c'est-à-dire le domaine de valeur d'une hypothèse. Cette notion d'hypothèse limitée à un ensemble avec des frontières est une première tentative épistémologique de nuancer la négation (parcellariser les hypothèses).
Les règles qui seront adjointes à celles des figures précédentes traitent de la quantification universelle et de la quantification existentielle. La quantification universelle est représentée dans la figure 4.
En préalable, on va supposer que x représente une variable de A, et c'est sur x que portera la quantification. Les autres lettres minuscules seront des occurrences de x, soit

libres — c'est-à-dire non déterminées par d'autres formules et elles seront décrites par des lettres minuscules de la fin de l'alphabet— soit des constantes (quand il s'agit par exemple des lettres a, b, c…).
Une autre façon d'exprimer A(x) est de dire que **A est une loi sur x**. La notion de loi décrit des propriétés régulières, à partir desquelles le système devra raisonner. Par exemple, la mortalité est une loi pouvant être applicable sur des organismes dits vivants. La loi va être considérée comme une hypothèse, et il faut proposer une axiomatique qui gère sa cohérence. La variable sur laquelle s'applique une loi est aussi appelable **fait** ou élément. Les faits observés et déterminés sont alors des constantes. Le fait «Paul Durand» de la loi Homme, si Paul Durand existe, est une valeur de la variable x dans Homme(x).

8) Si x est toujours lié dans Γ, $\dfrac{\Gamma \Rightarrow A(x)}{\Gamma \Rightarrow \forall x A(x)}$.
Qui signifie que si l'hypothèse A(x) est déductible du système, et qu'elle comprend une variable x qui est toujours liée dans ce système (qui est toujours déterminée dans et/ou par toutes les formules du système) alors l'hypothèse est déductible du système pour toutes les valeurs de sa variable. C'est la règle d'introduction de ∀, ou **règle de généralisation**.

9) Pour les occurrences t de x libres dans A(x), $\dfrac{\Gamma \Rightarrow \forall x A(x)}{\Gamma \Rightarrow A(t)}$.
Si un système permet de déduire qu'une hypothèse A(x) est toujours vérifiée quelles que soient les valeurs prises par sa variable, alors on admet que le système permet de déduire qu'elle est vérifiée pour des occurrences de cette variable qui sont libres dans cette loi. C'est la règle d'élimination de ∀, que l'on peut assimiler à une **règle de prédiction.**

10) En particulier : $\dfrac{\Gamma \Rightarrow \forall x A(x)}{\Gamma \Rightarrow A(x)}$ *et* $\dfrac{\Gamma \Rightarrow \forall x A(x)}{\Gamma \Rightarrow A(a)}$.
Si une hypothèse est toujours déductible d'un système quelle que soit la valeur de sa variable, alors elle est déductible du système in abstracto. C'est la **généralisation faits-loi**. De la même façon, toute valeur constante pour la variable d'une hypothèse déduite d'un système crée une hypothèse vérifiée de facto par le système, c'est **l'héritage loi-fait**.

Figure 4. Les axiomes concernant le quantificateur universel.

Commentaires

La quantification universelle, quand elle est en prémisse, introduit l'idée d'une propriété régulière. En revanche, quand elle est en conclusion, elle doit être vue comme une contrainte de nécessité, ou *condition nécessaire*. Ainsi, dans un cas, on «remarque» la ressemblance et on crée la classe d'équivalence des variables pour lesquelles l'hypothèse est une loi (hypothèse vraie), et dans l'autre, on signale que pour bénéficier du statut de loi, l'hypothèse A doit être nécessairement vérifiée sur le(s) domaine(s) de valeur de sa (ou ses) variable(s).

Les trois axiomes de la figure 4 gèrent en réalité l'**induction logique pure**. Pour les examiner, nous commencerons par le dernier. Si quels que soient les faits accessibles à un système cognitif, ces faits obéïssent à une loi vérifiée par ce système, alors on dira que cette loi est une règle du système, ou une connaissance, et ce, jusqu'à ce qu'elle soit falsifiée (et donc retirée). Les faits induisent la loi, *moyennant l'ensemble actuel des connaissances*. La deuxième partie de l'axiome 10 est classique, et signifie tout simplement que tout fait relevant de la loi la vérifie.
Exemple :
si, dans le système, on remarque que tous les humains vivants observés sont morts au bout d'un certain temps, alors on peut déduire que les humains sont mortels, en tant que loi, étant donné l'ensemble des connaissances du système observateur. Cela signifie que pour «Paul Durand», s'il est une valeur de x pour laquelle on a Homme(x), alors Mortel(«Paul Durand») est une hypothèse déductible du système.

L'axiome 9 est celui qui permet une prédiction sur les occurrences libres. En effet, à partir de la loi observée par le système, on va supposer que tout fait, même non encore observé, à partir du moment où il peut être un fait de cette loi (c'est-à-dire qu'elle est susceptible de s'appliquer sur lui) vérifie cette loi, pour le système.
Ainsi, cet axiome va permettre de dire que pour les hommes inconnus et non observés par le système, ce dernier peut «prédire» qu'ils seront mortels. De la même façon, le système pourra étendre la propriété de mortel à tout fait qui serait compatible avec les faits relevant de la loi de mortalité.

Enfin l'axiome 8 est celui qui permet justement de créer la loi A(x) comme loi, et non plus comme un fait lui-même observé. Si on s'aperçoit qu'à chaque fois que l'on donne une valeur à x qui est déjà initiée/proposée par une autre formule du système de connaissances, et que cette valeur fait que A(x) est vrai, alors on érige A en loi, c'est-à-dire on va créer la classe d'équivalence de tous les x pour lesquels A(x) est vrai. C'est là le début de la généralisation… ou le danger de l'induction, bien que ce danger réside tout autant dans l'axiome 9.

En effet, rappelons que ces axiomes supposent des propriétés sur x, et qu'il ne faut pas se lasser de les répéter. Il suppose aussi un système de connaissances qui est lui-même révisable. Ce qu'il faut dire c'est que,

- *étant donnés des faits observés et intégrés par un système,*
- *et étant donné l'état des connaissances d'un système,*

celui-ci s'autorise pour ces faits et par rapport à ces connaissances à attribuer momentanément une valeur de vérité à une hypothèse.

II.1.3 La quantification existentielle : le renforcement de la déduction prédictive

Les axiomes de la quantification existentielle dérivent en réalité des possibilités de l'axiome 9 de la figure 4. Ces axiomes sont décrits ci-après dans la figure 5. La quantification existentielle est d'une certaine façon complémentaire de la quantification universelle. C'est ce que nous tâcherons de montrer dans le paragraphe destiné aux commentaires.

11) Pour toutes les occurrences t libres de x dans A(x), $\dfrac{\Gamma \Rightarrow A(t)}{\Gamma \Rightarrow \exists x A(x)}$.

Si toutes les occurrences libres de la variable permettent de vérifier, à l'aide des connaissances du système, une hypothèse A(x), alors, le système permet de dire qu'il existe au moins une valeur pour laquelle l'hypothèse A(x) est vérifiée. (introduction du quantificateur existentiel, règle de l'**attente**).

12) De façon particulière, $\dfrac{\Gamma \Rightarrow A(x)}{\Gamma \Rightarrow \exists x A(x)}$ *et* $\dfrac{\Gamma \Rightarrow A(a)}{\Gamma \Rightarrow \exists x A(x)}$.

Si un système vérifie une hypothèse A(x), alors il la vérifie pour au moins une valeur de sa variable, et si un système vérifie une hypothèse quantifiée A(x) pour une constante *a,* cette constante peut être considérée comme une valeur de la variable *x*.

13) Si b est une constante indiviuelle n'intervenant dans aucune formule de Γ, ni dans $\exists x A(X)$, ni dans C, $\dfrac{\Gamma \Rightarrow \exists x A(X) \;\; et \;\; \Gamma, A(b) \Rightarrow C}{\Gamma \Rightarrow C}$.

Si un système stipule qu'une hypothèse peut être vérifiée pour une valeur possible de sa variable, et s'il existe une constante, qui n'est pas cette valeur possible, mais telle que de façon totalement indépendante des connaissances du système, l'hypothèse sur cette constante et le système impliquent une autre connaissance, alors cette dernière est satisfaite par le système (règle d'élimination du quantificateur, ou **induction-déduction**).

Figure 5. Axiomes relatifs au quantificateur existentiel.

Commentaires

Le quantificateur existentiel apparaît, en conclusion d'une règle, comme l'expression d'une *condition suffisante*. On remarquera par exemple que l'axiome 11 est totalement symétrique de l'axiome 9 (élimination de la quantification universelle). Il signifie que si une hypothèse avait valeur de loi pour un certain nombre de faits non pré-déterminés par le système, alors, il existerait en réalité au moins un fait observable par le système qui relève de cette loi. C'est une mesure logique pour asseoir la qualité de loi d'une hypothèse supposée être satisfaite par les connaissances du système. Or, il faut savoir qu'une hypothèse est censée être satisfaite tant qu'elle n'a pas été directement falsifiée, et sa falsification est gérée par les axiomes de la négation (figure 3).

Une hypothèse du système génère ainsi une attente de sa satisfaction. C'est ce qui permettra au système de mener des raisonnements abductifs, dont nous parlerons dans le prochain chapitre.

L'attente du système fait que tout fait observable par la suite va être l'objet d'une tentative de rangement par le système cognitif dans le domaine de valeur des variables des hypothèses qu'il possède.

Enfin, l'axiome 13 est de notre point de vue, probablement le plus important mais aussi le plus dangereux s'il est appliqué à des connaissances naturelles.

Voici comment il se décompose. Supposons qu'un système de connaissances implique logiquement qu'une hypothèse est **applicable** ou **possible**, c'est-à-dire qu'il existe une valeur de la variable pour laquelle l'ensemble des connaissances du système conclue en faveur de l'application de cette hypothèse. Supposons maintenant que l'on observe un cas d'application de cette hypothèse qui est parfaitement *nouveau* (non existant dans Γ ni dans A(x)) et *déterminé* (une constante individuelle) et que ce cas, associé à l'ensemble Γ, implique une nouvelle connaissance C. Alors on en déduit que c'est

l'ensemble du système de connaissances qui implique C, sans intervention de l'hypothèse A, et que par conséquent on pourra déduire C des connaissances du système.
En ajoutant C sous forme de proposition, quantifiée ou non, on a rajouté une loi au système, et ce, en testant l'application d'une hypothèse possible du système. Tout se passe donc comme si d'une hypothèse possible on était en mesure de déduire une proposition ayant force de loi. L'induction existentielle apparaît comme une condition effectivement suffisante pour tirer une conclusion.
Le danger de cet axiome n'est pas sur le plan des systèmes logiques formels tant que ceux-ci ne sont pas censés rendre compte de connaissances naturelles.
Exemple :
supposons que nous ayions un ensemble de connaissances tel qu'il satisfait l'idée qu'il existe une «variable» *x* dans l'univers telle que Martien(x) soit vrai (il existe quelque part un Martien, au moins un, dont on a une connaissance théorique). Si maintenant on découvre un individu jamais vu (n'appartenant pas au système de connaissances), et n'étant pas le Martien théorique accessible au système, mais découvert sur la planète Mars, alors on va raisonnablement supposer que cet individu, que l'on appellera Gloub par exemple, est un Martien. L'hypothèse était possible, là elle est instanciée par un fait totalement en dehors des domaines de valeurs de notre système cognitif, ce fait étant une *découverte*.
Penser que Gloub est un Martien, associé à l'ensemble des connaissances que nous avons, va nous conduire par exemple, au vu de l'analyse chimique des tissus de Gloub, à supputer que la planète Mars possède une atmosphère de méthane (proposition générale). L'axiome 13 s'appliquera alors effectivement moyennant que l'on ait quand même vérifié que Gloub n'est ni une constante ni une valeur possible de l'équation de la composition atmosphérique d'une planète[2].
Le problème pour appliquer cet axiome est :

- il faut connaître le domaine de valeur de la variable *x* de A(x) et *savoir quelle est l'occurrence de x qui est telle que A(x) est vrai*
- il faut pourvoir discerner la totale indépendance de la constante *b* vis-à-vis de l'ensemble hypothétique (connaissances + hypothèse A)
- il faut pouvoir asserter la totale indépendance de *b* quant à la conclusion C.

L'exemple du Martien est totalement anecdotique, mais l'on remarquera que l'axiome 13 s'il s'applique raisonnablement, ne peut être infalsifiable que si on a démontré la totale nouveauté de Gloub et le fait qu'aucune créature ne modifie l'atmosphère de sa planète. Est-ce vrai ? Il est difficile de répondre de façon évidente par oui ou par non. Il est donc difficile de faire de l'induction-déduction dans les conditions d'application de l'axiome 13.
Une autre interprétation possible de l'axiome 13 peut concerner, outre les faits nouveaux, *les faits qui servent de catalyseur*. Si on examine la conclusion (en dénominateur de la barre de fraction de l'axiome) on s'aperçoit que A n'est plus nécessaire pour déduire C du système Γ. Par conséquent, tout se passe comme si l'examen de A applicable sur une constante, dont on vérifie qu'elle n'apparaît nulle part dans C (c'est d'ailleurs bien le rôle des catalyseurs que de ne pas participer aux

[2]Chose qui pourra s'avérer difficile, car par exemple sur Terre les êtres humains modifient la composition atmosphérique de leur planète. Si Gloub se trouve dans le même cas, alors on ne pourra pas appliquer l'axiome.

réactions qu'ils provoquent), conduit le système à déduire une connaissance nouvelle ou indépendante, la proposition C. Une réflexion doit être menée concernant l'interprétation de l'axiome 13 comme expression de la catalyse. Cette interprétation, comme l'autre, sont hypothétiques. Le cloisonnement de la pratique scientifique a malheureusement conduit les logiciens à ne pas choisir la voie de l'interprétation *in vivo* de leurs axiomes.

II-2 La représentativité du système inductif logique

Le système inductif logique tel que nous l'avons montré précédemment fonctionne sur des hypothèses explicites et implicites qui sont beaucoup plus fortes que celles qui régissent la connaissance naturelle. L'exemple associé à l'axiome 13 (interprétation par fait nouveau) tend à inviter à une grande prudence. Mais si on regarde clairement le rapport existant entre les conditions d'un système logique et leur rapport au monde, en d'autre termes la pragmatique de la formalisation des connaissances, il est essentiel d'examiner l'ensemble des hypothèses à vérifier.

II. 2.1 Hypothèses explicites du système

Les hypothèses explicites générales sont données ci-après. L'énoncé de leurs limites par rapport au problème des connaissances naturelles suit l'énoncé de chaque hypothèse.

- Le système relève du calcul quantifié des propositions ou de la logique des prédicats du premier ordre pour lesquels on possède au moins un système déductif complet

  Les connaissances naturelles s'expriment souvent par des prédicats du premier ordre mais sans nuance. Or les nuances tendent à faire préférer soit des systèmes logiques non monotones soit des systèmes multi-valués. De plus, certaines connaissances naturelles sont telles que les prédicats peuvent recevoir en argument un prédicat. Elles relèvent pour partie au moins de l'ordre 2, ou 3 (nous pouvons montrer que c'est vrai dans le cas de connaissances linguistiques). Or le deuxième théorème de Gödel montre que l'ordre deux ne possède pas de système complet pour la déduction.

- Le système possède un ensemble fini de connaissances (propositions) qui sont des formules bien formées (relevant d'un langage et de ses règles de composition)

  Même sur un domaine donné de connaissances, on ne sait pas pour l'instant exprimer ces dernières par des formules bien formées du langage des prédicats du premier ordre.

- Le système possède les connecteurs de négation, d'implication logique, de disjonction et de conjonction logiques

  La sémantique des connecteurs logiques formels et celle des connecteurs de la logique naturelle sont divergentes. On voit en particulier l'ambiguité de l'implication logique, et la pauvreté de la négation.

- Le système possède deux quantificateurs, l'existentiel et l'universel

Le quantificateur universel exprime les termes «tout» et «chaque». L'existentiel dans un système logique suppose l'accessibilité des valeurs de variable. Dans le domaine des connaissances naturelles, l'accessibilité n'est pas forcément vérifiée.

- Le système sait à tout moment attribuer une valeur (vrai, faux) à ses connaissances (décidabilité)

Ce point est probablement la pierre d'achoppement de la représentation des connaissances naturelles, et il a été déjà largement traité par des chercheurs en Intelligence Artificielle comme Clancey ou Brachman.

- La révision du système en mode «hypothético-déductif» est gérée par les axiomes concernant l'introduction et rétraction des différents connecteurs et par le Modus Ponens (le Modus Tollens en est déductible).

Comme on l'a vu ces modes résolvent les problèmes de redondance et de contradiction flagrante, mais pas ceux de la réelle pertinence ou de l'adéquation au réel. En particulier la réfutation des causes peut conduire à la rétraction d'hypothèses par ailleurs utiles alors que l'adjonction d'hypothèse peut provoquer des conclusions incongrues. ***Le problème vient souvent de la non-dissociation des potentialités cognitives d'un système par rapport à son fonctionnement (raisonnement) sur un cas.***

Les hypothèses explicites concernant la quantification sont :

- L'introduction du quantificateur universel en conclusion se fait sur toutes les variables liées (la généralisation est limitée aux variables liées)

La généralisation chez les agents humains a tendance à être exhibée sur les occurrences libres (ou inconnues) de la variable. L'observation d'un ensemble de cas (petit, limité), conduit à poser la généralisation de l'hypothèse comme une loi du système.

- Si la généralité est stipulée, alors toute variable libre en hérite, et en particulier les cas observés (héritage) ou à observer (prédiction)

*Le problème dans les connaissances naturelles c'est que l'héritage peut parfois être circonscrit (préférence d'un système non monotone de type «défaut»). Les prédictions aussi peuvent être infirmées. En outre les connaissances naturelles relèvent souvent d'***une généralisation par majorité au lieu d'une généralisation par unanimité***. Cela fait que la connaissance et la circonscription des valeurs de variable satisfaisant l'hypothèse est souvent à redéfinir. Ce qui est vrai majoritairement sur un domaine, ne l'est pas toujours (toutes occurrences), ni obligatoirement (pour une occurrence quelconque émergente).*

- Si une hypothèse existe sur des occurrences libres de sa variable, alors elle est satisfaite par au moins une occurrence de cette variable, on s'attend donc à l'observer (attente, ou expectative).

L'attente ou l'expectative dans les connaissances naturelles tend à être suremployée et le quantificateur existentiel est alors transformé en quantificateur universel.

- Si une hypothèse est plausible, et qu'elle s'applique à un fait que l'on sait définir comme totalement nouveau et indépendant, alors ce que l'on peut en déduire vient s'ajouter aux connaissances du système.

Cet axiome est susceptible d'introduire des hypothèses on ne peut plus discutables (en terme d'adéquation au réel). Malheureusement ce sont celles qui permettent à l'induction de pouvoir se servir ensuite de tous les mécanismes déductifs grâce à l'implication. En particulier, on s'apercevra qu'un fait observé niant une proposition du système qui est une conclusion d'une règle, va mener le système à s'auto-réfuter, par application du Modus Ponens, ou à réfuter l'instanciation de la prémisse de la règle par Modus Tollens. La réfutation de la prémisse, si elle est généralisée ,est une erreur logique profonde, qui est malheureusement mal gérée par les agents humains.

Exemple d'un système s'auto-réfutant ou réfutant sa prémisse : ainsi supposons dans notre système de connaissance sur les moteurs de voiture l'ensemble fini d'hypothèses déjà admis :

les bougies sont encrassées ⊃ ¬(le moteur démarre)
les bougies sont vieilles ⊃ les bougies sont encrassées
les bougies sont encrassées⊃ Il faut changer les bougies

où l'on a : A= les bougies sont encrassées
B= le moteur démarre
C= les bougies sont vieilles
D = il faut changer les bougies
On réécrit alors le système Γ sous forme :

$$(1)\ A \supset \neg B$$
$$(2) C \supset A$$
$$(3)\ A \supset D$$

On admet que pour le système, il existe des valeurs x de B tels que B(x). Donc on a :
$\Gamma \Rightarrow \exists x, B(x)$
et il existe des valeurs y de C tels que C(y).
$\Gamma \Rightarrow \exists y, C(y)$
alors apparaît la voiture V de moteur m et de bougies b non répertoriée dans le système et telle que B(m) est vraie mais telle que C(b) est vrai aussi.
L'expression de l'ensemble du système augmenté des faits s'écrit :
$\Gamma, C(b), B(m)$
$\Gamma \Rightarrow \exists y, C(y)$ *et* $\Gamma, C(b) \Rightarrow A$ (par la règle 2) donc par l'axiome 13 on en déduit $\Gamma \Rightarrow A$
A est quantifiable par les variables z, l'axiome de généralisation montre que :

$\dfrac{\Gamma \Rightarrow A(z)}{\forall z A(z)}$ en particulier quand z prend la valeur b donc on a A(b). Or l'unification de la règle 2 avec la valeur b, permet celle de la règle (1)$A \supset \neg B$ avec les valeurs correspondantes pour A (i.e, b) et pour B (i.e, m) nous conduit à considérer, par Modus Ponens que si A(b) est vrai et si la règle (1) appartient au système alors la conclusion, ¬B(m) est satisfaite par Γ.
On a donc un système tel que :
$\Gamma, C(b), B(m) \Rightarrow B(m)$
$\Gamma, C(b) \Rightarrow \neg B(m)$
ce qui implique par l'axiome de l'élimination de la négation, que c'est la totalité du système Γ qui prend la valeur «faux». Cela signifie que le système se remet en cause, mais il ne sait pas qu'est-ce qui l'a invalidé autrement qu'en retraçant son raisonnement. Si le système était celui d'un agent humain, ce dernier remettrait en cause l'implication $A \supset \neg B$. Un système automatique n'en a pas directement les moyens, sauf si on le munit d'une structure de contrôle qui utilise le traçage du raisonnement pour faire un retour arrière.

Si pour ne pas invalider le système on décide d'appliquer le Modus Tollens, on s'aperçoit qu'effectivement :
$\Gamma, B(m) \Rightarrow \Gamma$ (adjonction d'une hypothèse)
$\Gamma \Rightarrow A(b) \supset \neg B(m)$ (unification de la règle avec les constantes)
par transitivité
$\Gamma, B(m) \Rightarrow A(b) \supset \neg B(m)$
règle du Modus Tollens

$$\frac{\Gamma, B(m) \Rightarrow A(b) \supset \neg B(m) \quad et \quad \Gamma, B(m) \Rightarrow B(m)}{\Gamma \Rightarrow \neg A(b)}$$

qui va permettre d'éliminer A(b), ce qui est une bonne chose parce que cela signifie que pour cette voiture V et ce moteur m et ces bougies b, même si les bougies sont vieilles, le système, pour se «préserver» ne va pas déduire qu'elles sont encrassées. Il cherchera à rétracter l'instanciation de sa connaissance A avec la valeur b plutôt que la totalité de A. Il ne remet pas en cause ses connaissances. Si le système est celui d'un agent humain, ce dernier, pour maintenir son harmonie cognitive, prétendra que les bougies sont propres sans chercher à le vérifier. Quant au système automatique, s'il s'agit d'un système sans procédure réflexive (évaluation), il ne s'embarrassera pas plus que cela de la remise en cause de ses connaissances.

L'anomalie logique sur l'induction apparaît lorsque l'on considère de manière abusive l'axiome de généralisation :

$$\frac{\Gamma \Rightarrow A \supset \neg B}{\Gamma \Rightarrow \forall x A(x) \supset \forall y \neg B(y)}$$

sans vérifier qu'il ne traite que des occurrences liées de x et de y pour A et B alors que l'axiome d'élimination de l'existentiel (induction déductive) ne traite que des constantes indépendantes sans occurrences dans A et dans B ni dans aucun membre de Γ. C'est l'erreur de la généralisation abusive à partir d'un contre-exemple. Supposer que parce que la voiture V démarre, l'encrassement des bougies de manière globale n'empêche pas le moteur de démarrer peut conduire cette fois-ci à perdre le bénéfice de la loi qui, elle, s'applique sur tous les précédents faits observés (ses variables connues et/ou liées).

II.2.2 Hypothèses implicites du système

Les hypothèses implicites relèvent surtout des caractéristiques de la quantification que l'on ne répètera jamais assez.

- Un système logique formel quantifié avec des variables suppose que l'on sache définir le domaine de valeur de ces variables, et que ces variables soient accessibles.

  C'est probablement le plus grand problème pour les connaissances naturelles. La définition correcte des domaines de valeur n'est possible que dans les domaines d'ores et déjà très formalisables, ou très simples.

- Il suppose aussi que l'on sache déterminer avec précision toutes les occurrences liées et toutes les occurrences libres d'une variable.

  Cela suppose que soit ces occurrences sont dénombrables soit que l'on a le moyen de les caractériser, et de maîtriser l'ensemble des connaissances du système. C'est un problème d'échelle. Un système avec quelque dizaines de propositions est encore dénombrable. Un système avc plusieurs dizaines de milliers de connaissances va échapper complètement au dénombrement. Au mieux, c'est l'explosion combinatoire. Au pire, c'est le pointage des valeurs inconnues. Nous n'avons en tant qu'agent qu'une vision parcellaire et incomplète de ce sur quoi peuvent s'appliquer les catégories (propositions, classes d'équivalence, lois, règles) que nous postulons si allègrement.

En conclusion, si l'induction logique formelle est un mécanisme parfait, son degré d'application aux connaissances humaines est relativement faible. Pis encore, une induction non prudemment exprimée risque de proposer des lois fausses voire des actions tout aussi erronées si le système inductif est actif. C'est pourquoi nous verrons dans le prochain paragraphe que l'induction expérimentale, qui est pourtant le mode épistémologique de base des sciences (de la matière à l'humain), est probablement un des mécanismes les plus intrinsèquement générateurs d'erreurs. Cependant, ce n'est pas une raison pour ne pas l'employer. Il faut savoir circonscrire l'inductivité des lois expérimentales.

II-3 Induction expérimentale

L'induction, vue cette fois-ci sous l'angle du raisonnement naturel, est le mécanisme de pensée par lequel les lois expérimentales sont constituées. L'induction se divise donc en deux grands groupes : la **catégorisation**, qui d'une part permet de créer des classes d'équivalence ou de pseudo-équivalence, et l'induction proprement dite qui met en jeu une implication.

II.3.1 Les faits et les lois

Sur un plan symbolique, l'induction fonctionne de la manière suivante :
on observe un certains nombre de faits complexes tels que :
$A(x_1) \supset B(y_1),\ A(x_2) \supset B(y_2),\dots,A(x_n) \supset B(y_n)$
Les objets de racine *x* sont des faits. On observe sur ces faits une propriété A(x). Les objets de racine *y* sont aussi des faits. On observe de même sur ces objets la récurrence d'une propriété B(y). Les faits, sur le plan expérimental, sont des observables (ou observés). Par exemple, on va observer *n* organismes vivants de notre planète. On observera de la même manière, si on a le temps, *n* morts correspondant à la fin de vie de ces *n* organismes. Sur un plan logique, ces *n* occurrences vont correspondre à des valeurs possible d'une variable qui prend ses valeurs dans le domaine des «organismes» susceptibles d'être vivants.
Le fait complexe, ou propositionnel, qui est observé est la régularité de l'implication entre les *x* et les *y*, c'est-à-dire, pour notre exemple, que les *n* organismes aboutissent invariablement à un état de mort.

L'induction expérimentale consiste à postuler, à partir de cette observation , que :

- il existe une relation d'équivalence entre les x_i telle que l'on peut catégoriser ces x_i et créer une classe d'équivalence $\overline{X}$ (à partir d'une relation d'équivalence R_A traduisant leur propriété régulière) les représentant qui, dans un formalisme propositionnel quantifié (ou prédicatif) se présentera sous la forme de la proposition quantifiée A(x), *x* prenant ses valeurs dans $\overline{X}$; par exemple, on pourra définir la relation d'équivalence «est-vivant-comme»[3] qui va donner lieu au domaine de définition de la proposition quantifiée « Est un organisme vivant (x)» ;

- de façon analogue, il existe une relation d'équivalence entre les y_j telle que l'on peut catégoriser ces y_j et créer une classe d'équivalence $\overline{Y}$ (à partir d'une relation d'équivalence $R_B$traduisant leur propriété régulière) les représentant qui, dans un formalisme propositionnel quantifié (ou prédicatif) se présentera sous la forme de la proposition quantifiée B(y), *y* prenant ses valeurs dans $\overline{Y}$; par exemple, on pourra définir la relation d'équivalence «est mortel-comme» qui va donner lieu à «Est mortel (y)» ;

- l'induction étant qu'il existe un morphisme entre l'implication factuelle observée dans $A(x_1) \supset B(y_1),\ A(x_2) \supset B(y_2),\dots,A(x_n) \supset B(y_n)$ et l'implication propositionnelle logique considérée dans $A(x) \supset B(y)$ qui est déductible de l'ensemble Γ des connaissances de l'esprit humain qui veut que

$$A(x_1) \supset B(y_1),\ A(x_2) \supset B(y_2),\dots,A(x_n) \supset B(y_n),\Gamma \Rightarrow \forall x \in \overline{X},\ \forall y \in \overline{Y}, A(x) \supset B(y)$$

Dans notre exemple, il se trouve que les *x* et les *y* sont confondus (ce sont les mêmes objets observés) et donc notre formule devient :

[3] on vérifie bien que c'est une relation réflexive (un organisme est vivant comme lui-même), symétrique, et transitive.

$Est-vivant(x_1) \supset Est-mortel(x_1), Est-vivant(x_2) \supset Est-mortel(x_2), \ldots,$

$Est-vivant(x_n) \supset Est-mortel(x_n), \Gamma \Rightarrow \forall x \in \overline{vivant_x}, \forall x \in \overline{mortel_x}$

$Est-vivant(x) \supset Mortel(x)$

Où $\overline{vivant_x}$ est la classe des vivants observés et $\overline{mortel_x}$ la classe des mêmes organismes vus comme étant mortels.
En Français, cette formule se traduit par :
Etant donné que les organismes vivants observés sont morts (mort observée)
Etant donné nos connaissances
nous pouvons dire que ,pour tout organisme observé, s'il est vivant alors il est mortel.

II.3.2 La généralisation

La formalisation logique aura tendance à raccourcir l'écriture :
$A(x_1) \supset B(y_1),\ A(x_2) \supset B(y_2), \ldots, A(x_n) \supset B(y_n), \Gamma \Rightarrow \forall x \in \overline{X},\ \forall y \in \overline{Y}, A(x) \supset B(y)$

en :
$A(x_1) \supset B(y_1),\ A(x_2) \supset B(y_2), \ldots, A(x_n) \supset B(y_n), \Gamma \Rightarrow \forall x \forall y \quad A(x) \supset B(y)$

Ce raccourci est effectivement valide dans les systèmes formels car aucun système formel n'envisage de dépasser les limites des domaines de valeur des classes d'équivalence et donc des domaines $\overline{X}$ et $\overline{Y}$. En revanche, le mode de pensée humain est tel que ces classes sont peu ou mal connues, ou la relation d'équivalence n'y est en fait qu'une simple relation de tolérance ou de d'équivalence limitée. Donc, les quantificateurs sont souvent utilisés dans leur dernière formule de façon à prédire pour des objets que l'on aurait envie d'assimiler à des x ou des y, l'application de la règle $A(x) \supset B(y)$.

En d'autres termes, pour notre exemple, on aurait envie de dire :
Etant donné que les organismes vivants observés sont morts (mort observée)
Etant donné nos connaissances
pour tout organisme, s'il est vivant, alors il est mortel.

Même un peu abusive pour ses conclusions, cette formulation reste cependant moins dangereuse que celle que l'on voit souvent apparaître et qui correspond exclusivement à la conclusion :

Tout organisme vivant est mortel.

La différence donc entre l'induction et l'induction abusive consiste en l'oubli (ou le rappel) de la prémisse ensembliste préalable à toute conclusion quantifiée, c'est-à-dire la partie :
$A(x_1) \supset B(y_1),\ A(x_2) \supset B(y_2), \ldots, A(x_n) \supset B(y_n), \Gamma$
et l'oubli ou la méconnaissance des domaines de validité :
$\forall x \in \overline{X},\ \forall y \in \overline{Y}$

Mais ce n'est pas la seule possibilité. L'économie généralisatrice va encore plus loin. L'axiome 10 qui traite de la partie «généralisation faits-loi» d'un système inductif formel aura tendance à montrer que :

$$\frac{\Gamma' \Rightarrow \forall x \;\; \forall y A(x) \supset B(y)}{\Gamma' \Rightarrow A(x) \supset B(y)}$$

Si Γ' est le nouvel ensemble de connaissances de l'esprit humain (enrichi des faits observés).

Si on exprime cette formule en langue, elle signifie que l'on aura tendance à intégrer l'idée non quantifiée (donc stabilisée) que :
un organisme vivant est mortel.

II.3.3 La propagation de la déduction dans des lois induites

L'intérêt de produire des lois induites est d'augmenter l'ensemble des connaissances de l'esprit humain par observation, mais aussi de renforcer cette augmentation par les propositions quantifiées que l'on pourra en déduire.
Ainsi l'induction est doublement productive :
$\Gamma, A(x) \supset B(y) \Rightarrow \Gamma$ (on obtient un surensemble des connaissances de départ avec une proposition quantifiée et non plus seulement des instances de cette proposition)
et on recherchera toutes les propositions C n'appartenant pas à Γ telles que :
$\Gamma, A(x) \supset B(y) \Rightarrow C$, de ce fait ces propositions neuves viendront augmenter l'ensemble des connaissances.
Tous les axiomes des systèmes hypothético-déductifs permettent de produire ces propositions C. En pratique un système à base de connaissances programmé dans un langage déclaratif a cet énorme avantage de pouvoir exhiber toutes les déductions C. Les systèmes naturels de connaissances, ceux de l'esprit humain en l'occurrence, produiront essentiellement les déductions les plus pertinentes.
Prenons par exemple la loi d'action et de réaction de Newton que l'on pourrait exprimer de façon intuitive comme ceci :
tout corps exerçant une poussée sur un autre corps reçoit en retour une poussée de la part de ce dernier égale à celle exercée par lui.
Un clou enfoncé dans un mur relève de cette loi, ainsi que la poussée que nous exerçons sur notre support terrestre et qui s'appelle notre poids (bien que celui-ci soit dépendant du coefficient d'accélération g, on verra comment la loi d'attraction universelle est une loi induite par suite d'un mécanisme de recherche des conditions de distance entre les différents corps).
A partir de cette loi, on peut retrouver la poussée d'Archimède qui, elle, est un sous-cas de cette loi lorsque l'un des deux corps est l'eau (un liquide, un fluide). Les déductions que l'on peut faire de la loi de Newton de l'action et de la réaction sont multiples. Là, nous parlons de la déduction pure, celle qui consiste à apparier la nouvelle loi avec l'ensemble des connaissances.
En revanche, lorsque la présence de la loi va inciter le système à acquérir de nouveaux faits (ceux par exemple qui ont mené Newton à considérer la loi fondamentale de la dynamique F = mγ, et par la suite, la loi d'attraction universelle), et à tenter de les classer dans ses connaissances, nous nous retrouverons face à un double système d'induction-abduction dont nous parlerons à la fin du chapitre III.

II.3.4 Limites et dangers de l'induction expérimentale

Reprenons l'exmple proposé dans le paragraphe II.3.2, qui aboutissait à la conclusion : *un organisme vivant est mortel.*
Tant que cette loi n'est pas infirmée par l'exemple, elle sera conservée.
Dans ces conditions, ce qui peut risquer d'arriver au premier contre-exemple possible c'est un ou plusieurs des cas suivants :

- remettre en cause la définition de vivant
- remettre en cause la définition de mortel
- remettre en cause la généralité de l'implication.

II.3.4.1.La logique des défauts et l'aide aux ontologies

Le problème qui est posé par les deux premiers points est un problème lié aux ontologies aristotéliciennes. C'est pourquoi ces ontologies sont un tel domaine de recherche et de publication pour l'Intelligence Artificielle. Comment définir la propriété régulière A sur le plan des connaissances naturelles ? Est-ce qu'on est vivant parce qu'on respire ? Cela exclut beaucoup d'organismes ne possédant pas d'appareil de traitement de l'air, comme de nombreux unicellulaires. Est-ce qu'on est vivant parce que l'on se multiplie ? Ou est-ce qu'on est vivant parce que l'on meurt ? Ce qui amène d'ailleurs à des définitions circulaires difficiles à maîtriser.
C'est pourquoi, la meilleure façon de bloquer les questions philosophiques sans fin que vont amener les limites de la définition des classes d'équivalence, c'est de proposer des ontologies opérationnelles et des classes de tolérance, au lieu des ontologies aristotéliciennes et les classes d'équivalence.
Une **relation de tolérance** est réflexive, symétrique mais non obligatoirement transitive alors que la relation d'équivalence l'est. On remarquera que l'on circonscrit d'ores et déjà la transitivité sans laquelle la déduction pure n'existe pas.
Une ontologie opérationnelle est définie par des propriétés majoritaires mais non unanimes. Par conséquent, des éléments d'une classe de tolérance peuvent très bien ne pas être liés par la relation de tolérance. Ainsi, si l'on dit :

les oiseaux volent en général

et que «voler-comme» est une relation de tolérance, tout comme la classe «oiseau» est définie par des propriétés uniquement relevant de relations de tolérance, alors il n'existe aucun problème pour définir le défaut suivant :
$\frac{oiseau(x) \supset vole(x)\ :\ autruche(x)}{\neg vole(x)}$ qui signifie que si la prémisse de la loi est vraie et la proposition après les deux points est aussi vraie, alors la conclusion de la loi est obligatoirement invalidée.
Une autruche donc ne cesse pas d'être un oiseau même si elle ne vole pas.
En définitive, ce que le symbole «:» décrit, c'est la notion de propriété majoritaire. Mais comme on le voit, on n'est plus dans la logique classique. De ce fait l'implication est obligatoirement restreinte par restriction de sa prémisse.

II.3.4.2. La circonscription de l'implication

Une autre technique consiste à ne pas remettre en cause les définitions des classes d'équivalence, mais à remettre en cause l'implication elle-même. Ce peut être les cas des

implications relevant du «bruit» généré par la non discrimination de l'opérateur $\supset$. On peut ainsi éliminer les implications absurdes du type décrit dans la section I. On peut aussi éliminer celles dont les prémisses sont fausses en appliquant le Modus Ponens qui est une façon plus restreinte d'écrire le théorème de déduction de la figure 1.
On peut enfin proposer une restriction pragmatique de l'opérateur $\supset$ en l'empêchant par exemple d'exprimer la causalité mais uniquement des relations dont on sait qu'elles possèdent un ordre total et qu'elles sont univoques.
Entre autres, n'employer l'opérateur $\supset$ que dans le sens de l'inclusion (large) de classes est une façon d'en restreindre la sémantique. C'est ce que font les systèmes à base de représentation «objet» en informatique. Le seul problème c'est que la sémantique causale et active qui disparaît fait disparaître avec elle le potentiel de raisonnement des différentes représentations. On n'a plus de diagnostic. C'est ce qui explique qu'il faut réécrire des systèmes d'inférence «au-dessus» des langages de programmation à objet.
Deuxième solution, n'employer $\supset$ que dans le cadre de l'antécédence chronologique. On peut produire des raisonnements valides mais ils sont uniquement temporels. C'est une des raisons de l'engouement de l'Intelligence Artificielle pour le raisonnement temporel ces dernières années. Les systèmes sont fiables logiquement d'une part, et ils représentent la réalité modélisée d'autre part.
L'ennui c'est que c'est l'ambiguité même de l'implication qui en fait toute sa richesse épistémologique. Dès lors, on se retrouve façe à un principe de dualité somme toutes assez répandu en sciences :
si on veut la plus grande adéquation au monde alors on perd la sécurité sur la déductivité de l'induction, et si on veut la plus grande rigueur formelle, alors on réduit obligatoirement la taille de son adéquation au monde (soit par adéquation à un monde restreint, soit par adéquation approximative).

C'est une autre façon d'exprimer la misère fort réelle du réductionnisme scientifique. Ce dernier n'est pas toujours le résultat d'un choix délibéré mais est souvent l'expression de la contrainte de ce principe de dualité, dont nous avons indiqué quelques dimensions logiques et objectives.

II.3.4.3. Les connaissances tempérées par la croyance

Une manière de détourner le problème a été de considérer que les connaissances étaient modulées :
- soit par un coefficient de confiance comme dans les logiques floues, et les règles assorties de probabilités
- soit par un opérateur de croyance estimant que la valeur à accorder à une proposition est en réalité une valeur de la croyance qu'a le système en la validité de la proposition (opérateurs *espérer*,*croire*, *penser*,*savoir*).

Il est clair que c'est une façon de circonscrire l'induction en en rendant directement le système cognisant émetteur responsable. Dès lors, on ne fonctionne plus avec un système de connaissances Γ, mais avec des couples.
Dans le cas d'une représentation modalisée, un couple (O,Γ) où Γ est l'ensemble des croyances d'un observateur O décrites par l'extérieur (dans le cas des opérateurs de croyances)

Dans le cas d'une représentation avec une fonction de confiance (fitness) par un couple ($f_{(O)}$, Γ) où f est une fonction dont les valeurs sont définies par l'observateur O lui-même s'adressant à l'extérieur, et telle que :
$\forall C \in \Gamma, \exists a \in [0,1] \subset \Re \,/\, f_{(o)}(C) = a$
Pour toute règle de l'ensemble de connaissances (de l'observateur O) il existe une probabilité (réelle) dénotant le seuil de confiance que l'observateur O attribue à cette connaissance.

La différence épistémologique est de taille : dans le premier cas c'est un système cognisant inconnu qui juge des connaissances de O. Dans le deuxième c'est l'observateur lui-même qui établit ses propres estimations sur ses connaissances.
Les deux systèmes sont limités sur le plan de l'adéquation épistémologique.
Quelle est la pertinence épistémologique d'une probabilité ? Ne confond-on pas la précision des réels avec la précision du jugement ? En d'autres termes, le modèle numérique est-il pertinent ?
Pour le second, quelle est la validité épistémologique d'un ensemble de connaissances/croyances attribué anonymement à un observateur ?

Il est clair qu'on ne peut pas répondre de manière générale à ce problème. Le seul cas me semble-t-il où l'on peut estimer sans trop se tromper la valeur relative du système cognitif, c'est lorsque l'on peut en mesurer le «rappel», c'est-à-dire lorsque nous possédons une représentation parfaite du domaine et que l'on évalue alors l'écart du système cognitif de l'observateur par rapport à cette représentation. Je laisse au lecteur le soin de juger d'une telle possibilité par rapport au problème qui le précoccupe.

III- Abduction

L'abduction est le mécanisme par lequel on rapproche le fait observé nouveau de faits connus ou de lois connues. L'abduction est un des raisonnements scientifiques les plus utilisés pour classifier des phénomènes, à partir d'un corpus constitué de lois (ou hypothèses cognitives). C'est un mécanisme qui ne fonctionne pas tout seul en général. Quand est un fait est reconnaissable, alors l'abduction est le principe par lequel on tente de reconnaître les propriété régulières de l'élément et de le classer comme membre d'une classe d'équivalence connue.
L'abduction est actuellement très utilisée en Intelligence Artificielle pour :

- augmenter les bases de faits classifiés
- pouvoir produire de nouveaux faits (déduits) en instanciant les lois déductibles de la loi concernée par le fait classé
- raisonner par analogie.

A propos du raisonnement par analogie, qui n'est pas un raisonnement au sens mathématique du terme, l'abduction est utilisée comme mécanisme de «forçage» et on en traitera dans la sous-section III.2. Parmi les promoteurs de cette utilisation on trouve Jerry Hobbs (pour la métaphore), Bipin Indurkhya, et bien d'autres.

III.1 L'unification

On posera les notations suivantes.
X, Y, Z… sont des formules prédicatives du premier ordre.
x,y,z… sont des instances respectives de X,Y,Z… avec des valeurs données pour leurs variables. En d'autres termes, les x_i, y_j,z_t sont des faits relevant des lois X,Y, Z.

Exemple :
si X est un prédicat avec deux arguments, que l'on notera *X(v,w)*, alors l'instance x_1 de X est le fait correspondant aux occurrences $v_1$de *v* et $w_1$de *w*,et l'on pourra noter sans problème l'égalité suivante : $x_1=X(v_1,w_1)$. Ainsi, si X signifie *est-père-de*, *v* représentera la personne parente, et *w* la personne descendante. La proposition factuelle suivante : «Jean est père de Matthieu» est une autre façon d'écrire *est-père-de* («Jean», «Matthieu»).

L'**abduction de type fait-classe** n'est autre que le procédé d'**unification**. Elle consiste à considérer tout fait nouveau et voir s'il relève d'un prédicat ou d'une formule prédicative connue.
Ainsi supposons que l'on recueille les propositions suivantes :
Pierre est père de Jacques.
Pierre est père de Marie
Einstein est le père de la relativité générale.
L'abduction fait-classe consiste à considérer que :
«Pierre» est une occurrence de v, et «Einstein» est une occurrence de v
«Jacques», «Marie» et « la relativité générale» sont des occurrences de w.

Bien sûr, on voit que les occurrences de w ne sont pas homogènes en type : comme l'homogénéïté de type est souvent requise pour les domaines de valeur, pour des raisons déductives évidentes, alors tout système comprenant des contraintes de type sur les valeurs de v et de w va rejeter le troisième fait.
Cependant, ce rejet apparaît comme violent, car quelque chose justifie la production de la proposition :

Einstein est le père de la relativité générale.

Cette paternité «étendue» est en réalité une autre façon de décrire le fait qu'Einstein est le concepteur de la relativité générale, comme Pierre a conçu Jacques et Marie, bien qu'il ne s'agisse pas des mêmes modes de conception.
Dans un réseau sémantique on écrira :
$est - père - de \subset est - concepteur - de$
qui signifie que la paternité est une sorte de capacité de concevoir. Sur le plan logique, si x est le père de y alors x est le concepteur de y. La table de contraintes est la suivante :

est-père-de	est-concepteur	contraintes
vrai	vrai	un père est forcément un concepteur
faux	vrai	un concepteur n'est pas forcément un père
faux	faux	un non concepteur ne peut pas être père

Table 5. Table des contraintes des deux prédicats

Le prédicat *est-concepteur* se comporte de telle façon que l'implication logique $est - père - de \supset est - concepteur - de$ est toujours vraie.
Dès lors, on s'aperçoit que ce qui sous-tend le problème de l'abduction, c'est l'existence d'une implication, c'est-à-dire d'une règle.
Le processus abductif des figures de style comme la métaphore et la métonymie consiste à remplacer le fait :

Einstein est le père de la relativité générale

qui serait une instance de *est-le-père-de* («Einstein», «la relativité générale»), et qui est mis en doute, par *est-le-concepteur-de* («Einstein», «la relativité générale»), qui peut-être vrai même si le premier est faux. Tout se passe donc comme si on avait :
soit t un fait observé. t ne relève pas de X, et $X \supset Y$, alors t ne peut-il être une instance de Y? C'est l'abduction par relâchement de contrainte dont nous reparlerons ultérieurement.

III.2 L'abduction par les effets

On s'intéressera plus particulièrement aux formules du type $X \supset Y$, c'est-à-dire aux lois qui sont des **règles** (comprenant une implication).
L'ensemble Γ de connaissances que l'on se donne comprend au départ :
$X \supset Y$
X, Y, Z…
et les x_i, y_j,z_t où les indices parcourent le domaine de valeur des occurrences de variables de chaque prédicat .

III.2.1 L'abduction restreinte

Une première forme d'abduction consiste, en examinant un fait nouveau qui est de structure implicative à tenter de le rapprocher de la règle $X \supset Y$ de l'ensemble des connaissances selon le schéma suivant.
Soit $t \supset u$ le fait observé. Si u peut être une instance de Y, et s'il n'existe pas au moins une formule prédicative X telle que $X \supset Y$ dans Γ et t est une instance de X, alors la règle $T \supset Y$ où T est une formule avec pour l'instant une occurrence unique t peut-elle être règle de Γ ?
Mathématiquement cela peut s'écrire :

$$\frac{\Gamma, t \supset u \Rightarrow \exists y_0 \;\; u = Y(y_0)}{\Gamma \Rightarrow \exists t_0 \;\; t = T(t_0),\;\; T(t_0) \supset Y(y_0)}$$

Commentaire
Concrètement cela signifie que lorsque les conclusions d'un fait observé peuvent relever d'une loi Y de l'ensemble des connaissances, alors les prémisses de ce fait vont être appariées avec les lois en prémisse des règles où Y est une conclusion. L'abduction de Peirce est une recherche des causes possibles.

L'algorithme le évident est le suivant :

```
pour tout fait t ⊃ u faire
    si ∃y0  u = Y(y0) alors
        pour tout X de Γ tel que X ⊃ Y faire
            si ∃x0  t = X(x0) est vrai alors
                        exhiber x0
                        valider X ⊃ Y
            finsi
        finpour
        si aucune règle de Γ n'est validée alors
                    créer T avec t = T(t0)
                    ajouter T(t0) ⊃ Y(y0) à Γ
        finsi
    finsi
finpour
```

Figure 6. Premier algorithme abductif.

On voit que la non attribution de t comme instance d'une classe possible va mener à la création de la classe T.
Le principal défaut de l'algorithme est que l'on risque de créer une classe nouvelle pour chaque fait qui ne relève pas d'une règle $X \supset Y$. Il faut donc tenter de circonscrire cela.

III.2.2 L'abduction inductive

Dans le cadre de l'agorithme précédent, la classe T créée n'est pas encore telle qu'elle puisse rejoindre le rang des causes X possibles. La rajouter comme classe est un processus qui n'est pas inductif au sens scientifique du terme, puisqu'en principe il faut observer plusieurs faits avant de produire une loi.
Pour améliorer l'algorithme, on va proposer de considérer les fait suivants :
pour tout nouveau fait, sa prémisse relève-t-elle d'une loi X telle que $X \supset Y$ ou d'une classe T telle que $T(t_0) \supset Y(y_0)$?
Si on peut montrer que cette prémisse est une instance de la classe T, alors on proposera de faire de l'induction et de donner momentanément un statut d'hypothèse à $T \supset Y$.

```
pour tout fait t ⊃ u faire
    si ∃y0  u = Y(y0) alors
        pour tout X de Γ tel que X ⊃ Y faire
            si ∃x0  t = X(x0) est vrai alors
                    exhiber x0
                    valider X ⊃ Y
            finsi
        finpour
        si aucune règle de Γ n'est validée alors
                pour tout T tel que ∃ti T(ti) ⊃ Y(yi) ∈ Γ faire
                    si ∃t0  t = T(t0) alors ajouter T ⊃ Y à Γ
                    finsi
                finpour
        sinon
                    créer T avec t = T(t0)
                    ajouter T(t0) ⊃ Y(y0) à Γ
        finsi
    finsi
finpour
```

Figure 7. Abduction inductive : algorithme d'ajout de lois partiellement vérifiées

On remarquera qu'avec cet algorithme une implication devient une règle du système de connaissance à partir de deux instances, où plus exactement, on crée des règles avec peu de faits. Cela peut servir dans le cas de bases de connaissances sur des phénomènes rares ou difficiles à observer, par opposition aux phénomènes fréquents qui vont provoquer la création inductive de loi par effet statistique. L'abduction est donc un procédé très intéressant de production d'hypothèses en particulier pour un système automatique parce que celui-ci va tester l'unification possible entre les classes et les faits observables. Les agents humains vont malheureusement avoir tendance a escamoter cette procédure d'unification, et on retombera dans le même travers qui peut endommager le raisonnement inductif.

III.2.3 La circonscription «naturelle» de la règle

La règle $T \supset Y$ est circonscrite aux valeurs d'unification. Cela signifie que si on n'arrive pas à trouver d'instance y de Y telle qu'il existe une instance t de T et $t \supset y$, cela ne pose pas de problème à la véracité de la loi.
Exemple :
Supposons que l'on ait observé une fois le fait : «un choc émotionnel violent a été subi par Dupont.» On constate une augmentation vertigineuse du taux de sucre dans le sang chez le patient Dupont.
L'analyse 135 du 12/03/98 de Dupont donne : taux-de-glucose-sanguin = 2,5 g/l.
Cela représente une instance possible d'une loi qui est *forme-de-diabète*.
On va donc écrire *forme-de-diabète*(«Dupont», 135) où l'on comprend que l'analyse de numéro 135 semble indiquer chez ce patient une forme de diabète.
L'algorithme de la figure 7 va permettre le raisonnement suivant :

- rechercher toutes les causes possibles de ce diabète parmi les symptômes présentés par le patient Dupont
- si les symptômes du patient (on ne trouve que le choc émotionnel violent) ne s'apparient avec aucune cause connue alors on va rechercher les autres cas de patients que l'on connaît déjà et qui auraient ainsi présenté le même tableau (ressemblance avec des situation connues)
- si effectivement on se rend compte que par le passé on a déjà observé un cas de choc émotionnel violent accompagné d'une hyperglycémie alors on peut se dire que au fond, bien que cela se produise rarement, un choc émotionnel violent peut être, à lui seul, la cause d'une hyperglycémie ;
- en revanche, si on s'aperçoit que l'on n'a jamais rencontré ce cas précédemment, alors on va ranger soigneusement le cas de monsieur Dupont dans sa mémoire et on se bornera à constater qu'en apparence le choc chez cet individu a précédé (a été cause de ?) son hyperglycémie.

Cela dit, dans un diagnostic, même si un médecin sait qu'un choc émotionnel violent peut être une cause de diabète, ce n'est pas la règle qu'il tentera de valider en premier. Le fonctionnement du raisonnement humain reste nettement à tendance majoritaire. Les règles rares induites par abduction sur les causes semblent avoir une priorité minimale. Dans un système artificiel en revanche, on peut tout à fait les essayer, puisqu'il suffit de vérifier que la prémisse observée est une instance de la classe en prémisse. Dans la mesure où l'algorithme de la figure 7 propose un ensemble de causes (et de règles) et non pas une règle unique, le raisonnement reste juste. Ainsi, supposons par exemple que l'on s'aperçoive que :

- Monsieur Dupont a reçu un choc émotionnel violent
- Monsieur Dupont a mangé beaucoup de gâteaux la veille de son analyse
- Monsieur Dupont est âgé de 60 ans

alors, on peut parfaitement se dire que sa forme de diabète est peut-être due à l'âge, peut-être à sa nourriture de la veille, et peut-être à ce choc. En soi, le raisonnement est juste, puisque les trois faits sont des instances de causes possibles (soit directement, soit par déduction).

L'action qu'entreprendra le médecin consistera à recommencer les analyses pour savoir si ce diabète est ancré ou s'il s'agissait d'un pic tout à fait circonstanciel. Cela dit, la cause que diagnostiquera le médecin n'est pas forcément la bonne, en raison des aléas de l'implication causale. C'est d'ailleurs pourquoi le diagnostic par les causes n'est généralement pas la façon la plus sûre d'arriver au résultat. Les praticiens préfèrent l'annulation des conclusions, c'est-à-dire quelles actions entreprendre pour que l'analyse 136 de Monsieur Dupont soit telle que *forme-de-diabète*(«Dupont», 136) soit faux. Recommencer tout simplement l'analyse est une des actions les plus évidentes.

III.3 L'abduction par les causes (ou abduction au sens de Peirce)

La forme symétrique de l'abduction par les effets consiste à tirer des conclusions à partir de faits où les causes tendent à montrer une certaine nouveauté.

C'est une abduction qui peut être tout aussi inductive que la précédente à partir du moment où l'on va passer des faits aux lois.

L'ensemble Γ de connaissances que l'on se donne comprend au départ :

$X \supset Y$

X, Y, Z…

et les x_i, y_j, z_t où les indices parcourent le domaine de valeur des occurrences de variables de chaque prédicat.

Soit $t \supset u$ le fait observé. Si t peut être une instance de X, et s'il n'existe pas au moins une formule prédicative Y telle que $X \supset Y$ dans Γ et u est une instance de Y, alors la règle $X \supset U$ où U est une formule avec pour l'instant une occurrence unique u peut-elle être règle de Γ ?
On arrive donc à apparier les causes. En revanche, les effets sont non encore classables ou inventoriables. On va tenter de les apparier avec des effets connus comme résultats des causes sélectionnées. Si cela n'est pas possible, alors on va considérer l'effet observé non pas comme faux, mais comme nouveau, et créer pour lui une classe, qui donnera lieu ou non à une vraie loi.
L'équivalent de l'algorithme de la figure 7 s'écrit pour l'abduction par les causes de la façon suivante.

```
pour tout fait t ⊃ u faire
        si ∃x0  t = X(x0) alors
                pour tout Y de Γ tel que X ⊃ Y faire
                        si ∃y0  u = Y(y0) est vrai alors
                                        exhiber y0
                                        valider X ⊃ Y
                        finsi
                finpour
                si aucune règle de Γ n'est validée alors
                                pour tout U tel que ∃ui  X(xi) ⊃ U(ui) ∈ Γ faire
                                        si ∃u0  u = U(u0) alors ajouter X ⊃ U à Γ
                                        finsi
                                finpour
                sinon
                                        créer U avec u = U(u0)
                                        ajouter X(x0) ⊃ U(u0) à Γ
                finsi
        finsi
finpour
```

Figure 8. Algorithme d'abduction inductive par les causes.

L'intérêt (en même temps que le danger) de l'augmentation des conclusions est la création de nouvelles déductions possibles. En effet, si la classe U existe dans le système de connaissances Γ et se trouve en prémisse d'une autre règle, alors le système satisfaisant $X \supset U \wedge U \supset Z$ va déduire $X \supset Z$.
Si l'inférence $X \supset U$ est abductive d'origine, l'inférence $X \supset Z$ est déductive.

III.4 L'abduction par relâchement de contraintes

On remarque que le fondement de l'abduction inductive, qui est au fond la seule abduction génératrice de lois (ce qui explique que pendant longtemps, l'abduction n'a pas été distinguée de l'induction) est la reconnaissance de la ressemblance entre un cas

nouveau observé et un cas existant préalablement inventorié. C'est l'analogie dite de **situation connue**. On arrive à produire du raisonnement sur le cas observé parce que l'on a déjà rencontré un cas relevant des mêmes classes (soit en cause, soit en effet) que l'on a étiqueté comme instance. On utilise donc au mieux la relation de ressemblance entre faits. C'est cette utilisation de la ressemblance qui est aussi à la base de l'analogie. Par conséquent, lorsque l'on observe un fait qui se comporte au moins partiellement comme un autre fait étiqueté, alors on relâche sur cet observable les contraintes et on extrapole la ressemblance, moyennant que cette extrapolation reste raisonnable. C'est comme cela que l'on peut raisonner par abduction pour décomposer les métaphores, pour raisonner à partir d'exemples (case based reasoning) et pour procéder à la résolution d'un problème en situation d'information incomplète.
On observe $t \supset u$. On possède une loi $X \supset Y$, et des lois de la forme $Y \supset Z$.
Premier cas : on a $\exists x_0 \quad t = X(x_0)$, et la sémantique de la proposition u voudrait que u soit une instance de Y.
Le problème est que, en raison du type des variables de Y u n'est pas instance de Y, c'est-à-dire, $\forall y_i \quad u \neq Y(y_i)$. L'algorithme est donné en figure 9 ci-après.

```
pour tout fait t ⊃ u faire
        si ∃x0  t = X(x0)  alors
                pour tout Y de Γ tel que X ⊃ Y faire
                   si ∀yi  u ≠ Y(yi) est vrai alors
                                répéter
                                        pour chaque Z de Γ tel que Y ⊃ Z
                                                si ∃z0  u ⊃ Z(z0) est vrai alors
                                                    valider u
                                                finsi
                                        finpour
                                jusqu'à ce que u soit validé ou il n'y ait plus de Z
                                finrépéter
                                si u est validé alors
                                        pour tout Z tel que Y ⊃ Z
                                                si ∃z0 tel que Z(z0) alors
                                                        ajouter u ⊃ Z(z0) à Γ
                                                finsi
                                                appliquer
                                (∃x0∃z0((X(x0) ⊃ u ∧ u ⊃ Z(z0)) ⊃ (X(x0) ⊃ Z(z0))))
                                        finpour
                                finsi
                        finsi
                   finpour
        finsi
finpour
```

Figure 9. Algorithme d'abduction par relâchement des contraintes.

Explication

pour tout Z de la forme $Y \supset Z$, on va essayer de vérifier $\exists z_0 \quad u \supset Z(z_0)$. Z, étant impliqué par Y, est moins contraint que lui puisque Z peut être vrai même quand Y est faux (c'est-à-dire quand on ne trouve pas d'instance de Y vérifiant la règle). Si on trouve au moins un Z_i tel que $\exists z_0 \ \ u \supset Z_i(z_0)$, alors on estimera que u est vrai, et pour tout Z, on va considérer que u permet de déduire une instance de Z et utiliser le fait que $(\exists x_0 \exists z_0 ((X(x_0) \supset u \wedge u \supset Z(z_0)) \supset (X(x_0) \supset Z(z_0))))$, c'est-à-dire le théorème de déduction.

Remarque : tel quel, cet algorithme n'est pas inductif, mais on peut le coupler avec l'algorithme d'abduction inductive de la figure 8.

Exemple : **compréhension de métaphore.**
Soit la phrase classique : *Marianne a des cheveux d'or*.
C'est une métaphore parce qu'il y a transfert de propriété depuis une source (l'or) vers une cible (les cheveux) alors que ces deux objets ne sont pas de même nature.
Ce que l'observe est l'implication :
$Possède(«Marianne», cheveux) \supset Etre-en-or(«cheveux»)$.
On va donc rechercher à réaliser des unifications, ou des abductions faits-classes pour raisonner sur cette implication observée.
On possède les prédicats :
Possède (x,y)
Humain (x)
Partie-de-corps (y)
Etre-en-or(z)
Chose (z)
Couleur (z,t)
et les règles suivantes :
1 $Humain(x) \wedge Partie-de-corps(y) \supset Possède(x,y)$
2 $Humain(x) \wedge Chose(y) \supset Possède(x,y)$
3 $Partie-de-corps(y) \supset Chose(y)$
4 $Etre-en-or(z) \supset Chose(z)$
5 $Chose(z) \supset Couleur(z,t)$
6 $Etre-en-or(z) \supset Couleur(z, jaune)$
L'unification permet de récupérer l'information suivante :
$Possède(«Marianne», cheveux)$ est déductible de $Humain(«Marianne»)$ et $Partie-de-corps(«cheveux»)$ et l'on a bien vérifié que Marianne est une instance d'humain et cheveux une partie du corps.(on supposera ici que Marianne est une personne, et pas la personnification de la République, sinon, on récupère une métaphore emboîtée et c'est difficile si on veut gérer l'exemple).
Le problème est que l'on devrait déduire de cela que $Etre-en-or(«cheveux»)$ est vrai, mais seules les choses peuvent être en or, or les cheveux sont une partie du corps. Il n'existe donc pas d'instanciation directe par $Etre-en-or(«cheveux»)$ d'une conclusion dont $Possède(«Marianne», cheveux)$ est la prémisse. On va donc relâcher une contrainte en essayant de voir quels sont les Z impliqués par $Etre-en-or(«cheveux»)$. Les règles que l'on trouve sont :
$Etre-en-or(«cheveux») \supset Chose(«cheveux»)$
$Etre-en-or(«cheveux») \supset Couleur(«cheveux», jaune)$
il faut donc vérifier qu'au moins une de ces instances est déductible par nos prémisses.

On montre ainsi $Chose(«cheveux»)$ est vraie en utilisant la règle $Partie-de-corps(y) \supset Chose(y)$ et en instanciant le prédicat *Partie-de-corps* avec la valeur «cheveux». La règle 3 de nos connaissances instanciée donne :
$Partie-de-corps(«cheveux») \supset Chose(«cheveux»)$
Donc la conclusion $Chose(«cheveux»)$ est vraie.
On vérifie incidemment que $Chose(«cheveux»)$ est vraie n'invalide pas le prédicat *Possède* observé (règle 2 de l'ensemble des connaissances).
A partir de là, on va considérer que u est vrai, et l'on va accepter les conclusions des implications où u est en prémisse. Comme u vaut $Etre-en-or(«cheveux»)$ alors $Couleur(«cheveux», jaune)$ est vrai aussi.
On va donc proposer comme conclusion la déduction qui est :
$Possède(«Marianne», cheveux) \supset Couleur(«cheveux», jaune)$.
On vérifie que cette conclusion n'est pas incompatible avec les autres règles.
Le Modus Ponens appliqué aux règles 3 et 5 permet de dire que :
$Chose(«cheveux») \supset Couleur(«cheveux», t)$
la conclusion obtenue peut être considérée comme une unification de t et par conséquent elle est préférable au prédicat partiellement unifié.

III.5 Les limites du mécanisme abductif

L'abduction peut être vue comme de l'unification à grande échelle, de l'unification réelle (fait-classe) à l'unification spéculative (instance de règle observée, règle stipulée). Le principe qui régit l'abduction est la reconnaissance d'une analogie, comme l'induction. Si l'induction introduit une relation de ressemblance (modélisable par la tolérance) entre plusieurs faits observés, l'abduction se fonde sur cette ressemblance pour proposer des lois, qui seront forcément restreintes, parce que l'abduction suppose quand même une vérification terme à terme des instances de classe.
Le mécanisme abductif est donc limité par les mêmes faits que le mécanisme inductif, soit :

- la qualité de la représentation ontologique
- les limites de l'implication
- le degré de confiance ou de représentativité des connaissances

III.5.1 La représentation ontologique et l'héritage multiple

Nous avions proposé pour ce premier point le remplacement d'une relation d'équivalence comme fondatrice de la classe par une relation de tolérance, moins rigide. Dans ce cadre, l'appariement fait-classe, principal mécanisme d'unification, peut amener à un appariement multiple, et l'on se retrouve dès lors confronté à l'ambiguïté. Le fait nouveau relève-t-il de la loi X, de la loi Y, ou de $X \cap Y$? Et si oui, comment gère-t-on les problèmes d'héritage multiple quand on fait de l'informatique ? Déduira-t-on ensuite à partir d'X, de Y, ou seulement à partir de règles dont les prémisses sont $X \wedge Y$? Tout dépend en effet comment sont X et Y. S'ils sont liés par une inclusion (ou une implication) alors le problème n'existe pas parce que l'héritage n'est pas multiple au sens de la programmation. Si en revanche il n'existe aucune relation logique entre X et Y, il sera difficile pour la mise en place de la déduction de choisir en $X \wedge Y$ et $X \vee Y$.

Par exemple si une occurrence factuelle, Marie Dupont, hérite des deux lois de tolérance «femme» et «employé de banque», notre connaissance et l'état du monde nous permettent de dire que, toutes choses étant égales par ailleurs, ce fait hérite de toutes les formules dans lesquelles au moins l'une de ces deux lois est présente.
Par ailleurs, si l'occurrence «Michel Dubois» s'unifiait avec «moins de treize ans» et «criminel» il ne pourra échapper à la maison de redressement et à la prison que parce que les deux conditions sont vérifiées. Dans un cadre donné, le cadre juridique par exemple, seule cette déduction est importante, et ce, même si l'occurrence peut générer d'autres instances de loi par déduction.
Le mécanisme abductif avec relation de tolérance a une incidence sur le modèle de représentation des connaissances. Il va invalider l'arborescence comme représentation de base et préférer celle de graphe. Le semi-treillis, et dans certains cas, le treillis, correspondent à de bonnes configurations. John Sowa plaidait pour le treillis, dans sa représentation de graphes conceptuels. Il n'avait cependant pas démontré formellement que le treillis était la seule structure possible.
Notre argumentation montre en revanche que, **si on se propose de définir une ontologie opérationnelle sur laquelle on souhaite faire de l'unification, alors l'arborescence n'est pas le bon modèle** pour la représenter, car un noeud nouveau peut être fils de plusieurs noeuds distincts.

III.5.2 Les limites de l'implication

L'abduction inductive engendre des lois partiellement vérifiées. Si la vérification, en induction, se fait sur plusieurs exemplaires, en revanche, en abduction, elle se fait sur peu d'éléments. Par conséquent, il faut se poser la question de ce que signifie :

- une loi $X \supset Y$ abduite
- si Z est créé à partir de l'observation d'une loi existante $X \supset Y$, et de $i \in [1,p], z_i \supset Y$, et l'on va en abduire $Z \supset Y$
- symétriquement, si c'est W qui est créé comme $X \supset W$ à partir de $X \supset Y$ et de l'observation que quelques instances de W sont des conclusions possibles pour X
- enfin que se passe-t-il si l'on a simultanément $Z \supset Y$ abduite, et $X \supset W$, ce qui aurait tendance à vouloir poser comme hypothèse la loi $Z \supset W$?

Pour le premier cas, cela peut rendre compte d' une inclusion de classe partielle. Or la circonscription de l'implication doit être représentée par la détermination exacte du domaine des valeurs d'instance de X et de Y. X et Y doivent être des classes tres restreintes, et si l'on veut permettre une telle loi, alors, justement, il faut revenir à une approche par classe d'équivalence plutot que par classe de tolérance. C'est le seul cas où les contraintes fortes de la représentation empêchent le système de dériver.

Pour les deuxième et troisième cas, la création des classes W et Z peut être légitime à condition qu'elles se restreignent aux valeurs observées, et ce, jusqu'à l'arrivée d'un nouvel observable qui soit les limite soit étend leur domaine de validité. L'implication est alors circonscrite d'un côté ou d'un autre par l'idée de classes dont la population est rare, par opposition à des classes du système de plus grande population. L'intérêt de l'implication qui a permis une abduction par les effets et par les cause est justement de produire de nouvelles abstractions dans le système, mais elles seront d'office révisables, non pas dans le sens de «vrai» ou de «faux» mais dans le sens de la dimension de leur domaine de valeur.

Enfin , c'est le dernier point qu'il faut considérer avec la plus grande prudence. On pourra se permettre d'examiner comme hypothèse à vérifier la proposition $Z \supset W$, avec une énonciation comme suit.
Si $X \supset Y$ appartient au système Γ relevant d' un couple (O,Γ), et si Z est induite par les effets (Z en prémisse) effet, et si W est induite par les causes (Z en conclusion), alors rechercher les $i, j \,/\, i \in [1,p], j \in [1,n],\ z_i \supset w_j$.
Cela signifie qu'il faudra tester le domaine de valeur d'une éventuelle proposition $Z \supset W$ Pour cela, une des manières de procéder est d'utiliser l'algorithme 9 dit par «relâchement de contraintes». S'il se trouve que seuls quelques éléments parmi les populations des classes induites se trouvent vérifier l'implication, alors il est préférable de définir la sous-classe limitée à ces éléments et la faire intervenir dans l'implication.
Ainsi on définirait $\begin{matrix} Z\min \supset Z,\ \ W\min \supset W \ \ \textit{tels que} \\ Z\min \supset W\min \end{matrix}$.

III.5.3 Génération des connaissances et attitude face aux faits nouveaux

L'abduction de manière générale, n'est pas pire que l'induction : l'abduction par analogie à une situation connue pourrait être moins nocive que l'induction généralisatrice. Il est bien plus dangereux, épistémologiquement, de créer une loi du type «tous les Athéniens sont menteurs» que de dire que «Castor est comme Pollux, il est donc menteur». Dans le premier cas, on détruit toute possibilité à tout fait instanciant la prémisse d'invalider la conclusion. Dans le deuxième on se trompe sur un fait; et un seulement.
L'abduction inductive hérite des mauvais côtés de l'induction. Si on dit :
«Castor est comme Pollux, il est menteur, ce Castor est d'ailleurs un Troyen, et je ne savais pas que les Troyens sont menteurs.»
Clairement, on outrepasse les conditions d'appplication de l'abduction inductive. Si en revanche, on disait :
«Castor est comme Pollux, il est menteur, ce Castor est d'ailleurs un Troyen,faisons donc une enquête pour savoir si les Troyens sont menteurs»
On se retrouve dans une condition où la proposition abduite est à tester par rapport au système et va nécessiter l'augmentation des connaissances de ce dernier avant de pouvoir être inclue dans Γ.

III.5.3.1 Intérêt fondamental de l'abduction dans la résolution des problèmes

L'abduction par les effets ou par les causes est probablement une des meilleures méthodes d'investigation en situation de données incomplètes sur un problème, c'est-à-dire lorsque le couple (O,Γ) se trouve dans l'incapacité, à partir de ce qu'il sait, et du mécanisme de déduction pure de résoudre un problème qui se pose à lui. Un problème est en général en occurrence unique : le couple (O,Γ) n'a pas forcément la possibilité de collectionner les faits multiples pour utiliser l'induction pure. Il est donc forcé de recourir à l'abduction, mais avec la restriction certaine que toutes les «connaissances»

qu'il va devoir acquérir pour résoudre le problème, ne sont pas des connaissances fiables.
Pour fiabiliser des connaissances, dans un système naturel comme dans un système artificiel, il faut du temps lorsqu'il s'agit de faits ou de problèmes inconnus. L'abduction se met en place de façon analogue à une mécanisme d'apprentissage. L'observateur va développer un appareil cognitif pour résoudre le problème, souvent en porte-à-faux par rapport à son propre édifice de certitude.
L'attitude raisonnable consistera ensuite à chercher de manière expérimentale à répéter le problème de telle manière à pouvoir observer un maximum d'instances des nouvelles classes créées, donc à provoquer de l'**induction expérimentale**.

III.5.3.2 L'abduction par relâchement de contrainte : le fondement de l'analogie

Les mathématiciens ont souvent signalé que l'analogie n'était pas un raisonnement au sens mathématique du terme, et ont toujours fait remarquer aux psychologues et informaticiens qui abusaient de cette dénomination que l'on ne pouvait appeler raisonnement un mode de développement des connaissances qui d'une part ne démontre rien, et d'autre part ne peut en chacune de ses étapes, être évalué à «vrai» ou à «faux». Sur un plan formel, les mathématiciens ont raison.
Si en revanche on considère que **l'analogie est une opération** (et non pas un raisonnement) **homothétique** (c'est-à-dire qui conserve les structures internes) qui s'applique sur un corpus de connaissances grâce à l'algorithme 9 décrit dans ce document, et dont l'objectif est de produire une connaissance hypothétique à partir de laquelle on peut appliquer une abduction inductive, alors ce processus apparaît comme formellement plus légitime.
On peut concilier l'observation psychologique et la rigueur logique par une telle description. Constatons que c'est l'informatique qui fournit ici la passerelle.

IV Une épistémologie de la praxis cognitive

Nous chercherons à résumer notre propos en posant des hypothèses qui doivent être soumises à instanciation et à la délimitation de leur domaine de validité.

1. Les principaux mécanismes de dérivation des connaissances sont la déduction, l'induction et l'abduction. Ils sont représentés dans les systèmes formels aussi bien que dans les systèmes naturels (humains) de cognition.

2. Les systèmes formels posent une axiomatique telle à ces mécanismes qu'elle en permet la décidabilité. Seules les formules mal formées seront indécidables. Les systèmes formels quantifiés permettent la révision de leurs connaissances de telle façon que, toute connaissances produite par un de ces mécanisme ou par une composition quelconque de ce mécanisme doit pouvoir être évaluée. Si elle ne l'est pas en vériconditionnalité, on appliquera des mesures d'évaluation de la confiance (logiques multivaluées, floues, etc.) En revanche, ces systèmes sont tels que leur mesures d'évaluation de la confiance ne sont elles-mêmes pas évaluables (théorème de Gödel).

3. Les systèmes naturels ont une dynamique cognitive qui va coupler en permanence les trois mécanismes principaux, et ce dans toute activité cognitive impliquant le raisonnement.

IV. 1 La recherche scientifique

En découverte scientifique, un des prototypes les plus usités pour la combinatoire est :

1.observation du fait nouveau (et confrontation au système de connaissances par **abduction fait-classe qui échoue**)
2.abduction inductive ;
3.recherche de **déduction** dans le système des connaissances
4.augmentation du système des connaissances sous forme de loi à étudier
5.recherche des observables appropriés
s'ils sont trouvés,
    6.1 **induction**
    6.2.attribution à la connaissance du statut de connaissance testée
    6.3.recherche d'augmentation du système des connaissances sous forme d'**abduction par relâchement de contraintes**
    6.4. retourner à l'étape 4
s'ils ne sont pas trouvés
    6.A. abandon ou retour à l'étape 3 (on cherche une autre déduction et s'il n'y en a pas on abandonne l'idée et on passe à autre chose)

C'est un script classique, et on observera que s'il est suivi à la lettre, il y a beaucoup de chance pour la loi obtenue ne soit pas directement invalidable. Une variante d'un tel scénario est de ne placer l'étape 3 qu'après l'étape 6.2. C'est souvent celle que le scientifique prudent va utiliser. L'attitude prudente est résumée dans le prototype suivant :

1.observation du fait nouveau (et confrontation au système de connaissances par **abduction fait-classe qui échoue**)
2.**abduction inductive ;**
3.augmentation du système des connaissances sous forme de loi à étudier
4.recherche des observables du domaine de valeur de l'abduction
s'ils sont trouvés,
5.1 **induction**
5.2.attribution à la connaissance du statut de connaissance testée
5.3.recherche de **déduction** dans le système des connaissances
5.4 recherche d'augmentation du système des connaissances sous forme d'**abduction par relâchement de contraintes**
5.5. retourner à l'étape 3
s'ils ne sont pas trouvés abandon

IV.2 L'apprentissage

En apprentissage, la combinatoire est quelque peu différente, mais à peine. La encore, il y a des variantes individuelles, culturelles dans les scenarii, mais voici un exemple de script qui a été observé par moi dans un groupe d'étudiants a priori peu alertes cognitivement (ce ne sont pas les plus brillants, ni même les plus motivés).

1.observation du fait nouveau qui est la connaissance à apprendre (et confrontation au système de connaissances)
2.1 augmentation du système des connaissances sous forme de fait
2.2 **abduction inductive ;** *("ah c'est la loi machin")*
2.3 **déduction** *("j'en déduis que c'est la loi qui s'applique dans tel cas")*
2.4. attribution à la connaissance du statut de connaissance testée
ou
2.A **abduction par relâchement de contraintes** *(ca ressemble à truc)*
2.B **abduction inductive ;** *("ah c'est la loi machin")*
2.C **déduction** *("j'en déduis que c'est la loi qui s'applique dans tel cas")*
2.D. attribution à la connaissance du statut de connaissance testée

ou
abandon

Je laisse au lecteur le soin de construire, par observation le scénario qui rend compte du comportement d'un ou plusieurs étudiants plus alertes ou plus critiques sur le plan cognitif. On peut pronostiquer un comportement proche de celui d'un scénario de recherche scientifique.

Pour conclure, il est clair qu'il existe a priori une infinité de prototypes de séquences, mais pour nous ce qui est évident, c'est que presque toujours, quand **le système cognitif est confronté à la nouveauté, il va tenter l'abduction en premier**.

Un système formel se comportera aussi de cette manière. Une bonne solution pour ne pas aller directement dans un diagnostic d'abandon est de forcer le système artificiel à appliquer un algorithme d'abduction inductive ou d'abduction par relâchement de contrainte.