Le théorème de Gödel et ses non-interprétations

Yann Ollivier

Introduction

Je donne ici plusieurs manières précises d'énoncer le théorème d'incomplétude de Gödel, ainsi qu'une esquisse de démonstration. J'essaierai ensuite de traiter rigoureusement quelques interprétations couramment avancées de ce théorème.

Pour lire ce texte, il n'est pas nécessaire d'avoir une formation spécifique en logique mathématique.

Table des matières

1 Le théorème d'incomplétude de Gödel

1.1 Très brefs rappels de logique

- On considère un système formel S, c'est-à-dire qu'on se donne un ensemble d'axiomes, et qu'on raisonne sur ces axiomes avec la logique usuelle.
- La propriété P est un *théorème* de S, ce que l'on écrit $S \vdash P$, si P est une conséquence des axiomes de S.
- Un *modèle* de S est un ensemble $\mathcal{M}$ dont les éléments satisfont les axiomes de S, ce que l'on note $\mathcal{M} \models S$.
- Une proposition P est dite *indécidable* dans S si ni P ni $\neg P$ (la négation de P) ne sont des théorèmes de S.

La première chose à noter est que l'indécidabilité n'a rien d'extraordinaire. Ainsi si l'on prend pour S les axiomes d'un groupe, la proposition qui exprime que le groupe n'a que trois éléments est indécidable, simplement parce qu'il existe des groupes à trois éléments et des groupes à plus de trois éléments ; autrement dit le système possède plusieurs modèles différents.

Plus précisément, un théorème de *complétude* de Gödel (qui malheureusement reçoit moins de publicité que son théorème d'incomplétude) précise que si une proposition est indécidable, alors il existe un modèle où elle est vraie et un modèle où elle est fausse. Autrement dit : si une proposition est vraie dans le système formel S (par *vraie* on entend vérifiée dans tous les modèles de S), alors elle est démontrable dans S.

1.2 Un énoncé du théorème d'incomplétude

Un énoncé du théorème de Gödel est le suivant : pour tout système formel S contenant le langage de l'arithmétique, il existe une proposition G indémontrable dans S (sauf si S est contradictoire, auquel cas il démontre n'importe quoi).

Il y a une restriction supplémentaire : le système doit être récursif, c'est-à-dire, en gros, qu'on doit pouvoir reconnaître les axiomes par un programme.

Je ne détaillerai pas ce que signifie « contenant le langage de l'arithmétique ». Par simplicité, on prendra par la suite, et définitivement, pour S la théorie usuelle des ensembles.

1.3 L'autoréférence gödelienne

L'idée maîtresse de la construction de Gödel est la suivante : il est tout à fait possible d'exprimer dans S ce qu'est la logique mathématique, ce qu'est une déduction, ce qu'est un modèle, etc. Partant, il est possible, dans le langage de S, d'exprimer une phrase qui dit « telle proposition est démontrable à partir de tels axiomes ».

Comment faire cela ? Rappelons que pour S nous prenons, par souci de simplicité, la théorie des ensembles usuelle. Il suffit alors de donner, dans S, des définitions de ce qu'est une suite de symboles, puis une suite de symboles syntaxiquement correcte (une proposition), de ce qu'est une déduction logique formelle (propriété d'une suite de propositions telle que chaque proposition découle des précédentes par applications d'une manipulation formelle correspondant à une déduction logique).

Par conséquent, comme S est assez puissant pour exprimer toute manipulation sur des suites de symboles, et qu'une preuve à partir d'axiomes est justement une certaine manipulation de suites de symboles, on peut parfaitement, pour tout système d'axiomes A (fini), produire une formule $Dem_A()$ de S telle que $S \vdash Dem_A(P)$ si et seulement si la proposition P du langage de A est démontrable à partir des axiomes A. (Il n'est pas nécessaire que le système d'axiomes A soit fini : il suffit qu'on puisse écrire une formule finie, ou un programme, capable de dire pour toute chaîne de symboles si oui ou non c'est un élément de A ; c'est le cas pour la théorie des ensembles usuelle.)

La formule $Dem_A(P)$ sera quelque chose comme : « il existe un entier n (longueur de la démonstration) et une suite de n suites de symboles (n propositions constituant la démonstration) telle que chacune de ces suites de symboles est soit un axiome de A, soit est obtenue par déduction de telle et telle manière à partir des précédentes, et telle que la dernière suite de symboles est exactement P ».

Remarquons bien le type des différents éléments : $Dem_A(P)$ est une formule du langage de S ; P est une proposition écrite dans le langage que nous avons utilisé pour faire faire à S de la logique (ce n'est pas une proposition de S), et A est un ensemble d'axiomes exprimés dans ce même langage.

On notera désormais ces niveaux d'imbrication par différents niveaux de guillemets : ainsi nous noterons sans guillemets (niveau 0) les propositions qui sont exprimées dans notre langage naturel : $1 + 1 = 2$. Nous noterons avec des guillemets “" (niveau 1) les propositions et objets du système formel

S : "$1+1=2$", que l'on devrait en fait noter "1" " + " "1" " = " "2". Nous noterons aussi S lui-même avec des guillemets. Nous savons, par exemple, que "S" $\vdash$ "$1+1=2$".

Enfin nous noterons avec des symboles $\lceil\rceil$ (niveau 2) les propositions et objets écrits dans le langage symbolique que nous avons développé pour faire faire de la logique à "S". Nous renoterons par " $\vdash$ " la proposition *Dem* ci-dessus, puisque c'est une proposition de "S" qui ressemble à notre $\vdash$.

Nous avons ainsi par exemple la formule suivante : "S" $\vdash$ "$\lceil A \to B, B \to C\rceil \vdash \lceil A \to C\rceil$".

Tout cela pour éviter des erreurs d'interprétation courantes dues au fait qu'on oublie qu'un système "S" ne peut pas parler *directement* des formules de "S". Autrement dit, une formule comme $\forall$"P", ... n'est pas une formule de niveau 1 (mais c'est une formule de niveau 0). Mais "S" peut très bien parler directement des formules de niveau 2 (c'est lui qui les a construites).

Maintenant, que signifie le fait que la formule " $\vdash$ " est bien l'analogue dans "S" de la notion de démonstration ? Cela signifie que notre procédé de construction de " $\vdash$ " garantit des propriétés comme

$$\left\{ \begin{array}{c} \forall\text{“}A\text{”}, \text{“}P\text{”}, (\text{“}S\text{”} \vdash \text{“}\lceil A\rceil \vdash \lceil P\rceil\text{”}) \Longleftrightarrow (\text{“}A\text{”} \vdash \text{“}P\text{”}) \\ \forall\text{“}A\text{”}, \text{“}P\text{”}, \text{“}Q\text{”}, \text{“}S\text{”} \vdash \text{“}\lceil A\rceil \vdash \lceil P \Rightarrow Q\rceil \Leftrightarrow \lceil A \text{ et } P\rceil \vdash \lceil Q\rceil\text{”} \end{array} \right. \tag{1}$$

et quelques autres du même acabit, que l'on démontrerait en explicitant la définition de " $\vdash$ ". Pour la première par exemple, à partir d'une démonstration symbolique du code de "P" à partir du code des axiomes "A", il me suffit simplement d'enlever les guillemets $\lceil\rceil$ pour obtenir une démonstration de "P" à partir de "A".

Attention cependant, ceci n'est vrai que sous certaines hypothèses sur "S" : à savoir, que tous les axiomes composant "S" soient vrais pour nous (sinon, on démontre n'importe quoi). Nous avons pris pour "S" la théorie des ensembles.

L'idée de Gödel consiste à faire étudier à "S" le système "S", plus exactement à regarder "$\lceil S\rceil$" dans "S". Maintenant, on va, par un procédé astucieux, fabriquer une formule G telle que G soit équivalente à « S ne démontre pas G ». Plus précisément, "G" sera équivalente dans "S" à la traduction de niveau 2 que " $\lceil S\rceil$ ne démontre pas $\lceil G\rceil$ " :

$$\text{“}G\text{”} = \text{“}\neg\,(\lceil S\rceil \vdash \lceil G\rceil)\text{”} \tag{2}$$

Cette phrase "G" est une n-ième version du paraxode d'Épiménide (« tout les Crétois sont des menteurs », ou encore « cette phrase est fausse »), qui a l'avantage de se prêter très bien à un traitement rigoureux, et qui est formellement inattaquable.

1.4 Construction de la phrase de Gödel

(Cette section plus technique peut être sautée en admettant l'existence de "G".)

La construction de "G" est comme suit. Remarquons d'abord qu'une proposition de niveau 1 est un simple objet au niveau 0 (une suite de symboles) ; de même une proposition au niveau 2 est en particulier un objet de niveau 1. On note d'abord "$g(x, y)$" la fonction suivante, qui prend deux objets de niveau 1 et renvoie un objet de niveau 1 :

- Soit "x" est un objet de niveau 1 qui est le code d'une proposition de niveau 2 à une variable : "x" = "$\lceil P() \rceil$" ("x" est une suite de symboles représentant une formule syntaxiquement correcte). On pose alors "$g(x, y)$" = "$\lceil P(\lceil y \rceil) \rceil$". À noter qu'alors, "$g(x, y)$" est aussi un objet de niveau 1 qui code une formule (sans variable libre) de niveau 2.
- Soit "x" est un objet de niveau 1 qui n'est pas le code d'une proposition de niveau 2 à une variable : on pose alors "$g(x, y)$" = "0" (le "0" de "S").

Il faut aussi ajouter à la définition de " $\vdash$ " la convention que "$x \vdash y$" vaut "faux" si "x" et "y" ne sont pas des symboles de niveau 1 codant des formules de niveau 2.

Ensuite, notons "$A(x)$" la formule "$\neg\,(\lceil S \rceil \vdash g(x, x))$". "$A$" est une formule de "$S$", donc on peut regarder son code "$\lceil A \rceil$" qui est une formule de niveau 2, donc un objet de "S". Notons donc enfin "G" = "$A(\lceil A \rceil)$".

En redéveloppant les définitions, on voit que

$$\begin{aligned} \text{“}G\text{”} &= \text{“}A(\lceil A \rceil)\text{”} \\ &= \text{“}\neg\,(\lceil S \rceil \vdash g(\lceil A \rceil, \lceil A \rceil))\text{”} \\ &= \text{“}\neg\,(\lceil S \rceil \vdash \lceil A(\langle A \rangle) \rceil)\text{”} \\ &= \text{“}\neg\,(\lceil S \rceil \vdash \lceil G \rceil)\text{”} \end{aligned}$$

où on a utilisé $\langle\rangle$ pour indiquer un objet de niveau 3.

Ce qui est bien la propriété annoncée.

1.5 Conséquences immédiates de l'autoréférence

L'existence de "G" a des conséquences immédiates très simples.

En particulier, "G" n'est pas démontrable dans "S" (sauf si "S" est contradictoire). En effet, supposons que "S" $\vdash$ "G". Alors, d'après notre construction de $\vdash$ (propriété 1), on a "S" $\vdash$ "$\lceil S \rceil \vdash \lceil G \rceil$". Mais par la définition de "G", dire que "S" $\vdash$ "G" revient à dire justement, que "S" *ne démontre pas* que "$\lceil S \rceil \vdash \lceil G \rceil$". Autrement dit, si "$S$" $\vdash$ "G", alors "S" démontre le faux, et est contradictoire.

De la même manière, on prouve que "$\neg G$" n'est pas non plus démontrable dans "S".

Une autre conséquence, qui constitue le second théorème d'incomplétude de Gödel, est que "S" ne peut pas prouver la formule "$\lceil S \rceil$ n'est pas contradictoire". La démonstration est la suivante : si "S" prouvait que "$\lceil S \rceil$" n'est pas contradictoire, alors "S" pourrait refaire, un niveau en-dessous, le raisonnement que nous venons d'écrire ci-dessus (où nous avons utilisé la non-contradiction de "S"), et prouver que "$\lceil S \rceil$" est soumis au théorème de Gödel ; autrement dit, "S" prouverait que "$\neg \lceil S \rceil \vdash \lceil G \rceil$"... mais ceci est précisément la phrase de Gödel, dont nous venons de prouver qu'elle n'était pas prouvable dans "S", si "S" était non-contradictoire.

2 Quelques non-interprétations

2.1 « Le théorème de Gödel prouve qu'il y a des vérités indémontrables. L'intuition humaine n'est pas formalisable. »

Pour la première affirmation, non : *si la phrase "G" de Gödel était vraie* (i.e. vérifiée dans tout modèle de "S"), *alors elle serait démontrable*, en vertu, précisément, du théorème de *complétude* du même Gödel. Il y a donc des modèles de "S" où "G" est fausse.

Cette croyance est suggérée par l'interprétation courante de "G" : « le système "S" ne peut pas prouver "G" », ce qu'on a précisément prouvé. Or "G" dit plutôt "le système $\lceil S \rceil$ ne peut pas prouver $\lceil G \rceil$". Quand on pense « le système "S" ne peut pas prouver "G" », on a affaire à une autre proposition, appelons-la G, et que l'on a bel et bien prouvée. Plus exactement on a prouvé que ou bien G était vraie ou bien "S" était contradictoire. Évidemment, "G" est la traductionn de G dans "S", mais ce qu'on attendrait que "S" prouve, s'il est aussi doué que nous, devrait être non pas "G" mais "G ou $\lceil S \rceil$ est contradictoire" ; et, précisément, cette phrase est prouvable dans "S" (i.e. en utilisant la non-contradiction de $\lceil S \rceil$, que "S" ne pouvait justement pas prouver, "S" peut arriver à prouver "G").

Ce qui est implicite dans le sentiment que l'intuition humaine dépasse les systèmes formels est le fait que nous-mêmes avons l'impression de travailler dans un système de raisonnement qui est la théorie des ensembles (en général), et qu'on a choisi le système "S" précisément pour coller à notre propre intuition : "S" est un candidat à notre formalisation. Par conséquent, *nous* avons réussi à prouver G alors que "S" ne peut pas prouver "G", d'où l'impression que l'intuition humaine n'est capturable par aucun système formel (la démonstration de Gödel s'applique à tout système formel assez puissant pour exprimer l'arithmétique, pas seulement à la théorie des ensembles).

Mais nous n'avons prouvé G que si "S" n'est pas contradictoire, ce dont nous ne savons rien pour le moment... Peut-être pouvons-nous le prouver.

Avant de raisonner sur ce que nous pourrions prouver, constatons que nous n'avons pas le droit de raisonner directement sur nos propres propositions et formules ; nous ne pouvons simplement pas dire « je peux prouver telle formule », nous ne pouvons pas écrire $\forall P, \ldots$, de la même manière que "S" ne peut pas écrire "$\forall$"P"$, \ldots$". Par contre, nous pouvons fort bien écrire $\forall$"P"$, \ldots$ Pour pouvoir analyser les raisonnements que nous pouvons mener, il convient justement de nous mettre à la place de "S", et, bien sûr, de remplacer "S" par $\lceil S \rceil$.

Supposons donc un instant que nous soyons un système formel "S" et regardons-le. On sait que si "S" admet que $\lceil S \rceil$ n'est pas contradictoire, alors "S" peut prouver "G" (ce que nous faisons depuis le début). Mais justement "S" ne peut pas prouver que $\lceil S \rceil$ n'est pas contradictoire... à moins que lui-même soit contradictoire. Donc, si nous sommes un système formel S, à moins d'être nous-mêmes contradictoires, nous ne pourrons pas prouver que "S" n'est pas contradictoire, et nous pouvons prouver seulement que « G ou "S" est contradictoire » : nous ne faisons pas mieux que "S" (forcément, si nous sommes un système formel). *Nous ne pouvons même pas prouver que nous ne sommes pas contradictoires* (sauf si nous sommes contradictoires).

Et nous ne pouvons même pas prouver que nous ne pouvons pas prouver que nous ne sommes pas contradictoires (sauf si nous sommes contradictoires). Etc.

Continuons à discuter l'affirmation « l'intuition humaine n'est pas formalisable ».

Justement, nous avons l'impression que nous, nous nous rendons compte des moments où nous rencontrons une phrase de type Gödel, et où pour continuer il faut supposer que nous ne sommes pas contradictoires. Eh bien, dans tout système S, on peut ajouter la connaissance que "S" est non contradictoire, connaissance que S pourra utiliser pour démontrer, par exemple, des phrases de Gödel portant sur "S". Mais ce faisant, par ajout d'un axiome, on a obtenu un système différent S', qui peut bien parler de la gödelisation de "S", mais pas de celle de "S'". Pour que S' puisse parler de sa gödelisation, il faudrait lui ajouter la non-contradiction de "S'", obtenant ainsi un second système S'', etc. *Il se peut fort bien que nous autres soyons modélisés par un système obtenu par un grand nombre d'itérations de ce processus*, et que nous ayons ainsi l'impression d'être un système S qui peut parler de la gödelisation de "S", alors que ce n'est vrai qu'à un grand ordre. Ceci peut expliquer notre impression que nous réussissons face à Gödel là où les systèmes formels échouent. [1]

Ces quelques niveaux de non-contradiction que nous nous accordons à nous-mêmes, nous les devons sans doute à la sélection naturelle ou à notre orgueil.

---

[1]Note pour les logiciens.

L'ordre auquel cette construction devrait être itérée est au moins 1 pour toute personne connaissant le théorème de Gödel. Il peut tout à fait être égal à un ordinal infini ; cela devrait être le cas pour toute personne ayant lu ce texte. Mais si cet ordre est égal à un certain ordinal α, on peut gödeliser à l'ordre $\alpha + 1$ pourvu que la réunion du système de départ et de toutes ses gödelisations jusqu'à l'ordre α soit un système d'axiomes récursif.

Il existe un plus petit ordinal γ où ce n'est plus le cas ; j'ignore quel il est (il est inférieur ou égal au plus petit ordinal non démonbrable pour des raisons évidentes de cardinalité).

Ceci est sans doute en rapport avec le fait qu'on ne peut pas nommer algorithmiquement tous les ordinaux dénombrables ; sinon, peut-être que cela pourrait permettre de reconnaître si une proposition donnée est une proposition de Gödel à un certain ordre. En tout cas, le théorème de Gödel prouve justement que l'ensemble des phrases de Gödel pour les ordinaux inférieurs à γ n'est pas calculable.

Il semble pourtant que, si “S” n’est pas contradictoire, la non-contradiction de $\lceil S \rceil$ est “vraie", puisque cela signifie la même chose ; auquel cas “G” serait “vraie" pour “S”. Mais cela repose sur notre assimilation implicite de « “S n’est pas contradictoire » et “$\lceil S \rceil$ n’est pas contradictoire", ou plus précisément sur notre assimilation de “ $\vdash$ ” et “$\lceil \vdash \rceil$”. On a bien nos formules de consistance 1, mais c’est insuffisant : dans un certain modèle de “S”, “$\lceil S \rceil$ n’est pas contradictoire" est bel et bien *fausse*.

Comment se fait-ce ? Il existe des modèles non-standard de “S”, des modèles, par exemple, où certains entiers sont considérés comme arbitrairement grands. Dans ces modèles, “$\lceil A \rceil \vdash \lceil P \rceil$” signifiera donc quelque chose comme : “ il existe une démonstration, éventuellement de longueur non standard, utilisant éventuellement des propositions de longueur non standard, de $\lceil P \rceil$ à partir de $\lceil A \rceil$". Ce n’est pas tout à fait la notion intuitive de démonstration que nous avons... L’assimilation entre “ $\vdash$ ” et $\vdash$, entre G et “G”, qui nous donne l’impression que puisque G est vraie, alors “G” devrait être vraie bien qu’indémontrable, n’est valable que dans *certains* modèles de “S”.

En ce sens, on peut avoir l’impression que c’est “S” qui capture mal notre intuition des entiers ou des ensembles. En tout état de cause, le théorème de Gödel prouve définitivement que tout système formel assez puissant (pour exprimer l’arithmétique) comporte deux modèles différents.

Si nous croyons que nous avons en tête un modèle intuitif unique, il ne peut donc pas être formalisé. Mais jusqu’ici, nous avons simplement vu que notre intuition comprend, outre les propriétés habituelles d’entiers et d’ensembles, quelques propositions de Gödel itérées à un certain ordre, qui nous font dire que pour les entiers intuitifs, il y a correspondance entre “G” et $\lceil G \rceil$. Notre modèle intuitif est non pas la version naïve des propriétés des entiers ou des ensembles, mais la même chose avec en plus l’axiome que “G” est vraie ; et aussi que la phrase de Gödel “G'” du système ainsi obtenue est vraie, etc., ceci itéré jusqu’à un rang assez élevé, comme ci-dessus, pour qu’on ne voie plus rien et qu’on ne puisse même plus repérer si telle proposition est une proposition qui affirme une vérité de Gödel. Notre modèle intuitif n’est pas unique, mais les branchements se situent à un niveau de gödelisation itérée très élevé (éventuellement égal à un ordinal infini). Et il est hors de portée logique, hors de portée de notre entendement fini, de parler de « tous » les niveaux de gödelisation.

Un cas particulier de « modèle standard » : si pour "S" on prend l'arithmétique, on sait, en effet, qu'en définissant $\mathbb{N}$ par la théorie des ensembles, le modèle $\mathbb{N}$ va vérifier le "G" de l'arithmétique. Cela veut simplement dire que la théorie des ensembles est plus précise sur l'arithmétique que les axiomes usuels (de Peano) que l'on prend pour les entiers, ce qui n'est pas une surprise (la théorie des ensembles est plus forte, elle contient en particulier un niveau au moins de gödelisation des entiers). Mais on ne sait pas faire la même chose pour la théorie des ensembles, vu qu'on ne sait pas en construire de modèle.

On peut aussi interpréter le « modèle standard » non pas comme celui qu'on a en tête sans jamais arriver à le décrire, mais comme celui du monde physique (en admettant qu'il y ait des nombres entiers « physiques » ou des ensembles « physiques »), de la réalité qui nous entoure. Dans ce cas, je ne vois pas quoi d'autre qu'un sentiment injustifié nous pousse à croire que G s'applique à celui-là. Alors, la phrase G serait une vérité indémontrable, mais elle porterait sur le monde physique, non sur une vérité logique accessible aux humains mais pas aux systèmes formels.

Ce n'est pas nouveau qu'il y a des propriétés de la réalité physique qui ne sont pas formellement démontrables : aucune ne l'est. Si nous utilisons tous les jours des propositions qui nous semblent vraies portant sur le monde extérieur, nous ne pouvons en fournir aucune démonstration rigoureuse ; mais nous possédons un grand nombre de connaissances sur le monde extérieur simplement par sélection naturelle, car le cerveau est conçu pour produire des résultats à peu près corrects sur la réalité qu'il doit traiter pour survivre.

Si donc "G" est "vraie" en un certain sens, c'est en un sens portant sur la réalité et non logiquement nécessaire. En tout cas rien ne semble incompatible avec le fait qu'un être humain soit intrinsèquement non formalisable.

2.2 « Le théorème de Gödel prouve que les hommes ne sont pas des machines. »

Non plus (ou du moins il faudra qu'on m'explique plus en détail).

Une traduction de la phrase de Gödel est « Untel ne peut pas prouver cette phrase sans se contredire ». Évidemment, Untel, voyant cette phrase, voit qu'elle est vraie — et qu'il ne peut pas le prouver. Pourquoi ne peut-il pas le prouver, simplement en disant « quel que soit le système S, S ne peut pas prouver cette phrase sans se contredire. Par conséquent, si je suis un système formel, je ne peux pas prouver cette phrase sans me contredire » ? Il peut le dire, mais cela dépend de l'hypothèse « je suis un système formel » — et alors il se contredit.

Nous avons donc bien prouvé — et c'est une vraie conséquence du théorème de Gödel — que je ne peux pas *prouver* que je suis un système formel ; ce qui ne veut pas dire que je n'en suis pas un (cela fera une vérité physique, et non logique, qui sera indémontrable) ; la situation est la même que pour notre non-contradiction.

De toute façon, la comparaison est assez mal choisie, puisqu'en fait, les tenants de la thèse de l'intelligence artificielle forte (en gros, un homme est une machine à neurones) ne veulent pas comparer un homme à un système formel, mais plutôt à un automate fini : personne n'a prétendu qu'un humain était un système formel. *Un humain est un objet qui à chaque instant se trouve dans un état mental ou dans un autre, pas un objet qui a des axiomes vrais et qui démontre des théorèmes* ; il n'explose pas s'il prononce deux phrases contradictoires, il se contente d'être dans tel ou tel état mental. Quand il voit un objet, il ne demande pas de preuve que cet objet est tel ; ce qu'il voit rappelle à son esprit les caractéristiques de l'objet ; ainsi pour G et "G". Les réflexions sur les limitations des systèmes formels sont donc à côté de la question.

Bien sûr certains comportements de cet automate consistent à prendre pour vrais dans la réalité les résultats qu'il obtient quand il choisit de simuler un certain système formel ; mais cette croyance est sans démonstration (on ne démontre rien rigoureusement sur la réalité extérieure).

Pourquoi ces automates ont-ils intégré de telles croyances ? Par sélection naturelle. La capacité d'un animal à croire qu'il est un système formel obéissant à certaines règles permet de déduire des choses sur lui-même et est certainement un pas important vers la conscience.

Conclusion : indécidabilité et paradoxes

Le plus petit nombre qui ne peut pas être défini par moins de dix-neuf mots en français

Discutons maintenant, à la lumière de tout cela, le paradoxe classique du plus petit nombre qui ne peut pas être défini par moins de dix-neuf mots en français (et que nous venons de définir, en français, par dix-huit mots).

Appelons donc S « le français ». Construisons une expression de S considérant les nombres que peut définir "S" : « $\inf\{n \in \mathbb{N}, \nexists$"$t$" terme de "$S$" de longueur $\leqslant K$, tel que "S" prouve que "$t = n$"$\}$ », expression dans laquelle K est bien sûr à remplacer par la longueur de cette même expression (qui, dans le détail, dépend en particulier de la manière dont on a défini inf, dont on a construit les guillemets, dont on a exprimé "S" etc.).

Maintenant, le raisonnement courant est : si cette expression avait pour valeur m dans S, cela serait évidemment contradictoire parce que dans S, m serait défini par moins de K mots, tandis que par définition sa définition aurait plus de K mots dans "S". Or nous (qui étudions S étudiant "S") savons de manière correcte que les deux doivent être les mêmes...

Effectivement, pour nous, S et "S" vont prouver les mêmes choses, mais il peut exister plusieurs modèles de S dans laquelle la valeur de m est différente, auquel cas il n'y aurait pas de preuve du tout... Si un modèle étudie un modèle différent, il n'y a donc pas de problème.

Le raisonnement « paradoxal » habituel prouve donc vraiment quelque chose : à savoir, que la valeur de m *n'est pas décidable* dans S, et qu'il existe des modèles où elle peut prendre des valeurs différentes (je ne sais pas lesquelles sont possibles). En effet, si la valeur de m était identique dans tous les modèles, alors S prouverait cette valeur, et donc le paradoxe s'appliquerait.

Encore une fois, cela ne prouve pas une quelconque « supériorité » du français sur S : S, de même que le français, s'aperçoit qu'il y a problème puisqu'il ne peut rien prouver. (Et on a vu que, de la même manière que le français peut dire « défini en français », S peut très bien dire "défini dans S".)

Le caractère normal de l'indécidabilité

L'on voit sur cet exemple que l'indécidabilité est là justement parce que l'interprétation de la proposition indécidable soulève un paradoxe.

Il en va de même de la phrase de Gödel ; si l'on suit précisément la construction, cette phrase est : « "est indémontrable si précédée d'elle-même entre guillemets" est indémontrable si précédée d'elle-même entre guillemets ». Soit exactement « cette phrase est indémontrable », phrase face à laquelle le lecteur doit d'abord constater qu'elle ne peut pas être fausse... bien qu'il ne puisse pas la démontrer ! Où l'on rejoint le « cette phrase est fausse ».

L'indécidabilité de la phrase de Gödel, ainsi que de la valeur du plus petit nombre dont la définition..., sont donc heureuses : au contraire, si S arrivait à démontrer une valeur de vérité précise pour « cette phrase est fausse », cela serait inquiétant ! Un système formel, en cette circonstance, réagit donc exactement de la manière attendue : quand on lui demande de démontrer un paradoxe, il n'y arrive pas, et c'est heureux.

Avec cette interprétation, l'on voit combien il est aberrant de dire que la phrase de Gödel est une « vérité inconnaissable »...