

Le paradoxe de Newcomb : testez votre philosophie !

Author : Bruno Jarrosson

Categories : [Philo Contemporaine](#)

Date : 10 avril 2014

En 1960, le Dr Newcomb, spécialiste de physique théorique aux Laboratoires des radiations de l'université de Californie à Livermore, énonça un nouveau paradoxe. Ce paradoxe, moins inquiétant que les paradoxes logiques qui ont miné les fondements de l'édifice mathématique, est cependant fort riche d'enseignements. En particulier, il trace une ligne de partage nette entre deux attitudes philosophiques opposées. Sous l'apparence d'un jeu d'esprit, le paradoxe de Newcomb offre donc un moyen clair d'attirer dans l'orbe de la philosophie les esprits qui lui sont allergiques. Ou indifférents, c'est-à-dire méta-allergiques.

Le paradoxe de Newcomb s'établit sur une communication avec un Être imaginaire qui possède la faculté de prévoir les choix humains avec une précision qui n'a jusqu'alors pas connu de démenti. Vous savez que cet Être a, par le passé, souvent prédit vos choix avec justesse et n'a jamais, pour autant que vous le sachiez, formulé de prévision incorrecte. Vous savez aussi que cet Être a souvent prédit avec justesse les choix des autres personnes dans la situation particulière ci-dessous décrite.

Voici le jeu que l'Être vous propose : il vous montre deux boîtes et vous explique que la boîte 1 contient mille euros tandis que la boîte 2 contient soit un million d'euros soit rien du tout selon sa propre décision. Deux alternatives vous sont offertes : prendre ce qui se trouve dans les deux boîtes ou bien ne prendre que ce qui se trouve dans la seconde boîte. L'Être vous explique qu'il a arrangé les choses comme suit : si vous choisissez la première alternative et prenez le contenu des deux boîtes, l'Être qui l'a prévu laissera la boîte 2 vide ; vous ne recevrez donc que mille euros. Si vous décidez de prendre seulement la boîte 2, l'Être qui l'a prévu y met alors un million d'euros.

Voici la séquence des événements : l'Être opère sa prévision puis, selon ce qu'il a prédit de votre choix, met ou bien ne met pas le million d'euros dans la boîte 2, puis il vous communique le principe du jeu puis vous faites votre choix.

Cette situation étant posée, il ne reste qu'à faire le choix, le contenu de la boîte 2 ou les contenus des boîtes 1 et 2. Ce choix constitue bien entendu l'objet de la discussion.

Raisonnement 1 : je puis avoir une confiance presque totale dans la faculté qu'a l'Être de prédire mon choix. En conséquence, si je décide de prendre les deux boîtes, l'Être l'aura prévu presque à coup sûr. Il aura donc laissé vide la boîte 2. Si je décide au contraire de ne prendre que le contenu

de la deuxième boîte, l'Être aura presque certainement prévu ce choix et y aura mis le million d'euros. Il est par conséquent raisonnable de ne choisir que la boîte 2. Où est donc le problème ?

Raisonnement 2 : au moment où je dois effectuer mon choix, les deux boîtes ont déjà été remplies. Il est donc clair que le contenu de la boîte 2 plus le contenu de la boîte 1 est supérieur ou égal au contenu de la boîte 2 seul. Il est donc tout aussi clair que le gain obtenu en prenant les contenus des deux boîtes sera supérieur ou égal au gain obtenu en prenant le seul contenu de la boîte 2. Il est par conséquent indiscutable qu'il faut prendre les contenus des deux boîtes. Où est donc le problème ? On vous l'a dit et répété.

Mal raisonné, rétorquent aussitôt les partisans de la première option : c'est précisément ce raisonnement que l'Être a presque certainement prévu. Par conséquent, il aura laissé vide la boîte 2.

Vous ne saisissez pas, répliquent à leur tour les défenseurs du second point de vue : l'Être a achevé sa prévision et le million d'euros est ou n'est pas déjà dans la boîte. Peu importe ce que vous décidez, car l'argent était ou n'était pas là depuis une heure, un jour ou une semaine avant que vous ne prissiez votre décision. Cette décision ne le fera pas disparaître s'il est déjà dans la boîte, ni apparaître s'il n'y est pas. Vous commettez l'erreur de croire ici en une sorte de causalité *a posteriori*, à savoir que votre décision est susceptible de faire apparaître ou disparaître, selon le cas, le million d'euros. Mais l'argent a été placé dans la boîte ou gardé à l'extérieur avant que vous ne vous décidiez. Et dans les deux cas, il serait trop stupide de ne prendre que la boîte 2, car si la boîte 2 est pleine, pourquoi négliger les mille euros de la boîte 1 ? Au pire, si la boîte 2 se révèle vide, vous serez content de gagner au moins mille euros.

On voit que le débat se cristallise sur ce qui sépare le **réalisme** de l'**idéalisme**.

La force du raisonnement 2, c'est d'être parfaitement réaliste. Il postule en effet qu'il existe une réalité indépendante de mon esprit et indépendante des décisions que je prendrai pour l'observer. Cette réalité, c'est le contenu – inobservé – des deux boîtes. Comme *a priori* il ne peut y avoir de causalité qui remonte le temps, le raisonnement semble imparable. Dans une optique réaliste tout au moins.

Les tenants du raisonnement 1 n'ont pas toujours conscience de leur rupture avec le réalisme. Pourtant ils réagissent bien de façon idéaliste. Ils croient les yeux fermés en l'hypothèse de départ : l'Être sait prédire avec exactitude ; le résultat n'est pas commandé par la décision du sujet mais par la prévision que l'Être en aura faite. Dans la mesure où cette prévision est exacte, tout se passe comme si elle commandait la décision du sujet. L'idée de l'Être commande – positivement – les faits, d'où la qualification d'idéalisme philosophique.

Ceci étant vu, quelques remarques s'imposent. Tout d'abord, l'histoire proposée par Newcomb pour semer le doute dans les esprits est elle-même idéaliste car la supposition de l'existence d'un

Être presque omniscient est manifestement irréaliste.

Ensuite, on peut penser que l'Être se demandera si le sujet est réaliste ou idéaliste. Le sujet qui en a conscience est engagé dans une méta réflexion : l'Être me prend-il pour une réaliste ou un idéaliste ? Je sais qu'il sait que je sais, etc. Jeu de miroirs insoluble et inévitable. Avant de se demander si je suis idéaliste ou réaliste, l'Être lui-même doit se poser une méta question : le sujet a-t-il deviné que l'Être se posera cette question ? Cette question renvoie aussitôt à une méta méta question, etc. L'Être n'a donc pas de méthode rationnelle pour connaître la décision du sujet. Ce qui renforce l'aspect idéaliste du paradoxe en augmentant l'omniscience de l'Être.

Autre point important, celui de la preuve. Comme indiqué plus haut, il ne peut être prouvé en aucune façon que l'Être ne se trompera pas. Aucun moyen de prouver que je ne puisse le « doubler ». La décision se fonde donc sur un acte de foi ; décider, c'est faire un pari. Ceci nous rappelle que toute décision relève d'un acte de foi car on n'a jamais la preuve, au moment où on l'engage, qu'elle atteindra son but. Et un acte de foi, comme le remarquait Pascal, est un pari.

La différence entre les raisonnements 1 et 2 peut également se comprendre à partir d'acceptions différentes du vocabulaire. Dans la phrase : « Si Bruno est le père de Gabriel, alors Gabriel est le fils de Bruno », « si... alors... » renvoie à une relation logique atemporelle entre les deux termes. Mais dans la proposition : « Si vous appuyez sur ce bouton alors la sonnerie retentira », le « si... alors... » indique une relation causale, et toutes les relations causales sont temporelles au sens où il s'écoule nécessairement un laps de temps, aussi minime soit-il, entre la cause et l'effet.

Or il se pourrait fort bien que les gens qui défendent le raisonnement 1 (ne prendre que la boîte 2) fondent leur raisonnement sur le sens atemporel de la relation « si... alors... » : « Si je décide de ne prendre que le contenu de la boîte 2, alors la boîte contient un million d'euros ». Les partisans du raisonnement 2 (prendre le contenu des deux boîtes) paraissent au contraire raisonner sur la base de la relation « si... alors... » causale et temporelle : « Si l'Être a déjà fait sa prévision, alors, selon ce qu'elle est, il a ou n'a pas mis le million de la boîte 2 et dans les deux cas je gagne mille euros de plus si je prends le contenu des deux boîtes ». Cette deuxième position se fonde sur la séquence causale et temporelle : prédiction de l'Être, mise de l'argent ou pas, choix du joueur. Dans cette perspective, le choix du joueur intervient à la fois après la prévision de l'Être et sa décision : il ne peut exercer aucune influence *a posteriori* sur ce qui a eu lieu avant lui.

Que mettons-nous derrière les mots ? Chacun les entend à sa façon car « l'homme est la mesure de toute chose ». Dans notre façon de comprendre les mots, dans notre choix entre les deux boîtes ou seulement la boîte 2 s'expriment des options philosophiques que parfois nous ignorons nous-mêmes.