



HAL
open science

Le dilemme de Parsifal

Léo Gerville-Réache

► **To cite this version:**

| Léo Gerville-Réache. Le dilemme de Parsifal. 2015. hal-01134102

HAL Id: hal-01134102

<https://hal.science/hal-01134102>

Preprint submitted on 22 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le dilemme de Parsifal

Léo Gerville-Réache - IMB - UMR 5251 - Université de Bordeaux

Résumé : Le dilemme de Parsifal est original car il oppose deux joueurs qui peuvent appartenir à deux catégories : les Parangons ou les Paresseux. Cette spécificité du jeu ne rentre pas dans les exemples standards de la théorie des jeux non-coopératifs à deux joueurs (dilemme du prisonnier, poule mouillée, mille-pattes ou autre...). Aussi, la plupart des résultats sur les équilibres de Nash et les stratégies mixtes ne sont plus valables. Comment se sortir de ce dilemme qui nous emmène jusqu'à la question de la pertinence de l'hypothèse de la connaissance commune de la rationalité, ou encore les limites des raisonnements conditionnels asymétriques?

L'histoire ¹...

"Sur la lointaine planète Par cohabitent paisiblement deux peuples très étranges : en effet, un milliard d'individus se répartissent équitablement en Parangons et en Paresseux. Tous les Parangons sont des modèles de rationalité, des artistes, des techniciens, des scientifiques et des philosophes curieux de tout, de fins calculateurs capables avec le temps de résoudre des problèmes difficiles, qui agissent toujours en vue de leur bien et avec de bonnes chances de succès. A l'opposé, tous les Paresseux sont des êtres qui, bien qu'amoureux de la vie, doués de raison et modestement cultivés, bien que malmenés parfois par le destin, réfléchissent en tous domaines le moins possible, trouvant simplement leur satisfaction dans l'obéissance et dans l'imitation. Les habitants de Par n'ont pas de goûts ou de préférences qui les distingueraient et forgeraient leurs personnalités : les Paresseux suivent tant bien que mal les courants et les progrès imposés par leurs voisins parangons, lesquels cherchent le beau objectif et l'utile absolu.

D'après les résultats des expériences effectuées par le vieux professeur parangon Paradoxe, connu de toute la communauté parangonne, si on demande à un Paresseux de participer à un jeu qui consiste à choisir une lettre de l'alphabet, il choisira une lettre, complètement au hasard et de façon imprévisible, indifférent aux conséquences de son choix même si un raisonnement lui permettrait de comprendre, par exemple, que seule une voyelle le fait gagner. Cependant, il choisira inmanquablement une voyelle à la condition qu'on lui explique très clairement qu'il ne gagne à ce jeu que s'il choisit une voyelle.

Un jour, Parsifal, jeune Parangon dynamique, est convoqué par le professeur Paradoxe. Celui-ci explique que, pour mener sa nouvelle expérience, il a choisi au hasard, avec équité totale, deux habitants de Par, en bonne santé et ni trop jeunes ni trop vieux : Parsifal est l'un des deux, l'autre individu se trouve ailleurs sur la planète. Paradoxe poursuit: chacun de leur côté, les deux heureux élus doivent choisir un des trois nombres 1, 2 ou 3, et seulement un de ces trois-là, et le noter sur une feuille. Si la somme des deux nombres est impaire, les participants repartiront chacun avec cent parpaings (la monnaie locale), sinon ils repartiront bredouilles. Placé devant sa feuille, Parsifal hésite. Il questionne en vain le vieux professeur, qui lui dit qu'aucun des deux participants ne doit savoir si son compagnon est un Parangon ou un Paresseux. Parsifal doit maintenant compter sur lui-même. Peut-il se donner plus d'une chance sur deux de gagner ? Que doit-il écrire sur sa feuille ?"

¹ Mes remerciements sincères à Laurent Delabre, l'auteur de cette très belle énigme.

Formalisation du jeu...

Un Paresseux choisira chacun des trois chiffres avec équiprobabilité. Un Parangon cherchera à maximiser sa probabilité de gagner. Le jeu comporte quatre combinaisons gagnantes (1;2), (2;1) (2;3) et (3;2) et cinq combinaisons perdantes (1;3), (3;1), (1;1), (2;2) et (3;3).

| | 1 | 2 | 3 |
|---|-----|-----|-----|
| 1 | 0;0 | 1;1 | 0;0 |
| 2 | 1;1 | 0;0 | 1;1 |
| 3 | 0;0 | 1;1 | 0;0 |

Table1 : Matrice des gains (100 parpaings=1)

La spécificité de ce jeu est que Parsifal peut se retrouver soit face à un Paresseux dont la stratégie est connue, soit face à un Parangon dont la stratégie est de maximiser sa propre probabilité de gagner. Il est également important de noter que le jeu n'est pas à somme nulle. En effet, les intérêts des deux joueurs sont totalement dépendants (i.e. lorsque l'un gagne, l'autre gagne aussi). Enfin, la matrice des gains est symétrique.

Pour Parsifal, si l'autre est un Paresseux, les choses sont claires. Parsifal doit choisir le chiffre 2 et chacun aura 2 chances sur 3 de gagner. Si l'autre est un Parangon, c'est plus délicat. Parsifal sait qu'il doit, autant que possible, ne pas donner le même chiffre que son partenaire. Aussi, tout raisonnement qui conduirait à une stratégie pure (jouer un chiffre choisi de façon déterministe), produirait l'égalité des nombres inscrits par les deux Parangons et rendrait la somme perdante. Aussi, Parsifal perdrait systématiquement face à un Parangon. Au global, la probabilité de gagner de Parsifal (dans le cas où il choisirait de jouer le chiffre 2) serait alors de 1/3.

Tout Parangon sait qu'il ferait déjà mieux en jouant l'un des trois chiffres avec équiprobabilité (la probabilité de gagner serait alors au global de 4/9). C'est donc une stratégie mixte que Parsifal se doit de chercher.

La question principale est de savoir si Parsifal doit supposer que tout Parangon conclura à la même stratégie mixte que lui ou pas. D'une manière générale, si l'on note P_2 la probabilité que Parsifal joue le chiffre 2 et P_2' la probabilité que tout autre Parangon joue le chiffre 2, la probabilité de gagner de Parsifal face à un Parangon vaut $P_2*(1-P_2')+(1-P_2)*P_2'$ (la probabilité d'inscrire 1 ou 3 étant supposée égale par symétrie de ces deux nombres dans ce jeu). La probabilité de gagner de Parsifal face à un Paresseux vaut simplement $1/3+1/3*P_2$. Comme la probabilité, pour Parsifal, de jouer face à un Paresseux est égale à celle de jouer face à un Parangon (vu qu'un milliard d'habitants vivent sur Par, on néglige le fait que Parsifal est un Parangon), la probabilité de gagner de Parsifal vaut :

$$P(G) = 1/2*(1/3+1/3*P_2)+1/2*(P_2*(1-P_2')+(1-P_2)*P_2') = 1/2*(1/3+4/3*P_2+P_2'-2*P_2*P_2')$$

Parangons identiquement rationnels...

L'histoire précise, " *Les habitants de Par n'ont pas de goûts ou de préférences qui les distingueraient et forgeraient leurs personnalités*". On est alors en droit d'étudier ce qui passerait si les stratégies des Parangons étaient nécessairement identiques (i.e. $P_2=P_2'$).

Dans ce cas, Parsifal a une probabilité de gagner :

$$P(G) = 1/2*(1/3+1/3*P_2) + 1/2*(2*P_2*(1-P_2)) = 1/6+7/6*P_2-P_2^2.$$

Le maximum est atteint pour $P_1=P_3=5/24$ et $P_2=7/12$. La probabilité de gagner de Parsifal est alors de $P(G)=1/2+1/144$ (voir figure1).

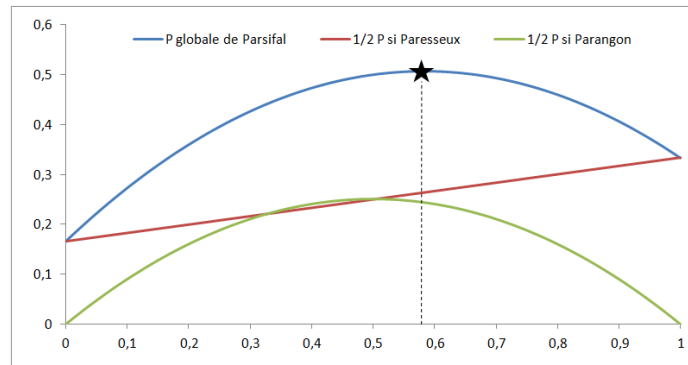


Figure 1 : Les options de Parsifal face à un Paresseux ou un Parangon

Tout semble clair :

- Un Paresseux jouera $P_1=P_2=P_3=1/3$. Il aura une probabilité de gagner face à un Paresseux de $4/9$, une probabilité de gagner face à un Parangon de $19/36$, et au global, une probabilité de gagner de $1/2-1/72$.
- Un Parangon jouera $P_1=P_3=5/24$ et $P_2=5/12$. Il aura une probabilité de gagner face à un Paresseux de $19/36$, une probabilité de gagner face à un Parangon de $35/72$, et au global, une probabilité de gagner de $1/2+1/144$.

Le doute de Parsifal...

Pourtant, un doute naît de la poursuite du raisonnement de Parsifal: "*Si, moi Parsifal, je décide de jouer le chiffre 2, j'augmente un peu mes chances de gagner, "opposé" que je serais, soit à un Parangon "classique" choisissant 2 avec probabilité 7/12, soit à un Paresseux*".

En effet, Parsifal choisissant de jouer $P_2=1$, se retrouverait dans la situation suivante :

- Face à un Paresseux, il aura une probabilité de gagner de $2/3$.
- Face à un Parangon (classique), il aura une probabilité de gagner de $5/12$.

Aussi, la probabilité de gagner de Parsifal sera au global de $1/2+1/24$. Cette probabilité est supérieure à celle obtenue initialement ($1/2+1/144$). Il semble donc que Parsifal ait finalement intérêt à jouer $P_2=1$.

La probabilité de gagner de Parsifal, en fonction de P_2 , vaut alors $P(G)=1/2*(1/3+1/3*P_2)+1/2*(5/12*P_2)$.

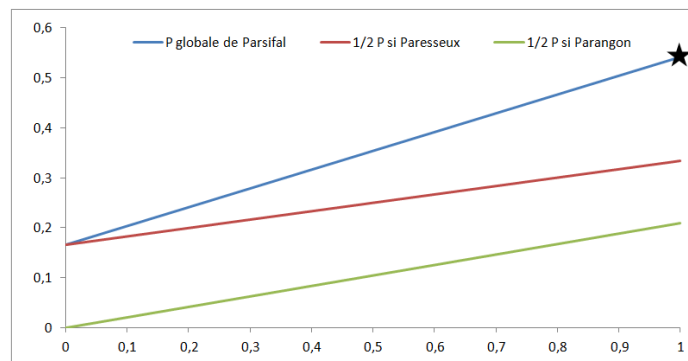


Figure 2 : Les options de Parsifal face à un Paresseux ou un Parangon "classique"

Cependant, ce doute de Parsifal peut naturellement naître dans l'esprit de tout Parangon. En effet, un Parangon "classique" qui serait face à Parsifal se retrouverait avec une probabilité de gagner de $1/2-1/12 = 5/12$ au lieu des $1/2-1/72 = 35/72$ qu'il serait en droit d'attendre (face à un Parangon

"classique"). Aussi, il est clair que le raisonnement tenu par Parsifal sera possiblement tenu par n'importe quel Parangon. Chaque Parangon pourrait alors jouer alors $P_2=1$.

Mais dans ces conditions, les probabilités de gagner de chaque Parangon (et donc de Parsifal) se trouveraient fortement bouleversées :

- Face à un Paresseux, Parsifal aura une probabilité de gagner de $2/3$.
- Face à un Parangon, Parsifal aura une probabilité de gagner de 0 .

Aussi, la probabilité de gagner d'un Parangon sera au global de seulement $1/3$!

Avec ce raisonnement, ce sont les Paresseux qui profiteraient de la situation. En effet, pour un Paresseux :

- Face à un Paresseux, il aura une probabilité de gagner de $4/9$.
- Face à un Parangon ($P_2=1$), il aura une probabilité de gagner de $2/3$.

Au global, la probabilité de gagner d'un Paresseux devient $1/2+1/18$ (contre $1/2-1/72$ initialement).

Parsifal, comprenant alors que les Parangons pourraient jouer $P_2'=1$, se révisé. La probabilité de gagner de Parsifal, en fonction de P_2 , face à un Paresseux ou un Parangon " $P_2'=1$ " vaut alors au global $P(G)=1/2*(1/3+1/3*P_2) + 1/2*(1-P_2)$.

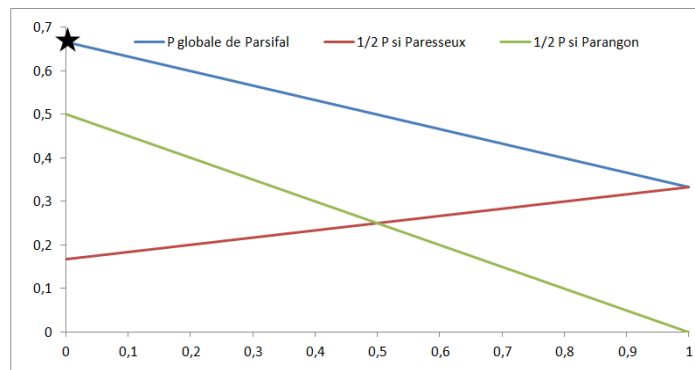


Figure 3 : Les options de Parsifal face à un Paresseux ou un Parangon $P_2'=1$

Si les Parangons jouent $P_2'=1$, Parsifal a clairement intérêt à jouer $P_2=0$. La probabilité de gagner face à un Paresseux sera alors de $1/3$. Elle sera de 1 face à un Parangon. Sa probabilité globale de gagner vaut alors $1/2+1/6$. Mais Parsifal, se devant de supposer que les autres Parangons pourraient jouer également $P_2'=0$, en conclut que sa probabilité de gagner, en fonction de P_2 sera alors $P(G)=1/2*(1/3+1/3*P_2) + 1/2*P_2$.

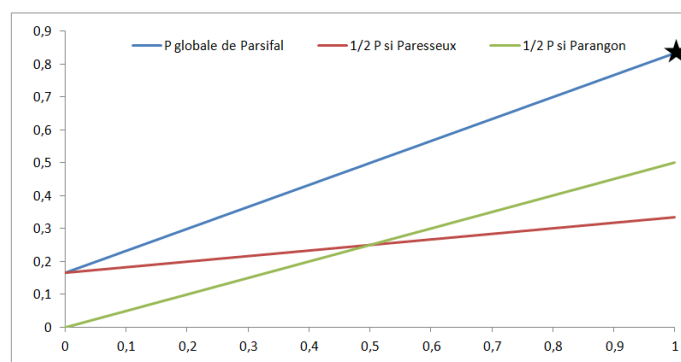


Figure 4 : Les options de Parsifal face à un Paresseux ou un Parangon $P_2'=0$

La probabilité de gagner est maximale pour $P_2=1$ et cette probabilité vaut $1/2+1/3$. Nous sommes ici dans le cas le plus favorable pour Parsifal. En effet, il joue $P_2=1$ face, soit face à un Paresseux, soit face à un Parangon jouant $P_2'=0$. La probabilité de gagner pour Parsifal est à son maximum possible du jeu : $P(G)=1/2+1/3=5/6$ (figure 4).

Aussi, chaque Parangon aurait intérêt à jouer 2 avec l'espoir que son adversaire, s'il devait être également un Parangon, jouerait tout sauf 2. On comprend ici tout le paradoxe de la situation. Est-il concevable que deux Parangons puissent mettre en œuvre des stratégies différentes?

En effet, rien ne les distingue et bien que les raisonnements analysent des éventualités asymétriques, il semble qu'une telle solution ne pourra être réalisée. Aussi, même si les raisonnements asymétriques sont théoriquement tenables, ils ne peuvent en aucun cas produire l'issue asymétrique qu'ils envisagent.

A moins que chaque Parangon finisse par décider de jouer au Paresseux. En effet, si le Parangon poursuit le raisonnement, il conclura qu'il a intérêt à faire le Paresseux. En effet, sa probabilité de gagner serait finalement de $4/9$ au lieu de $1/3$. La planète Par serait alors peuplée de vrais et faux Paresseux, chacun ayant une probabilité de gagner de $4/9$ (voir figure 5).

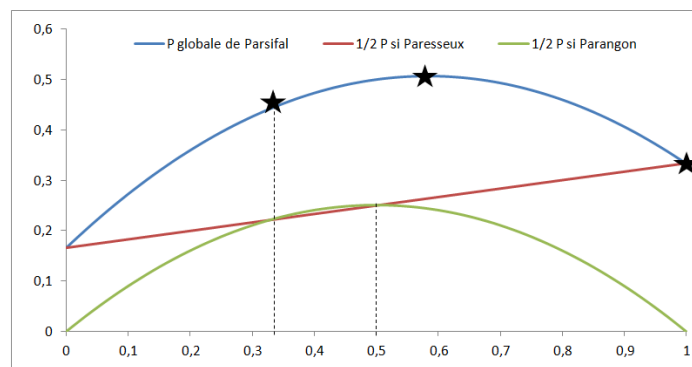


Figure 5 : Les différentes stratégies de Parsifal en cas de symétrie des stratégies des Parangons

Pourtant, Parsifal poursuivant son raisonnement se dirait que dans un monde peuplé uniquement de Paresseux (vrais et faux), aura intérêt à jouer "2" car sa probabilité de gagner sera alors de $2/3$. Voilà Parsifal revenu au point précédent et doit reprendre son raisonnement le faisant osciller à l'infini entre $P_2=7/12$, $P_2=1$ et $P_2=1/3$.

Comment est-il possible qu'un Parangon ne puisse se sortir d'un tel dilemme? Le raisonnement de Parsifal contient-il une erreur?

Une issue au Dilemme?

L'histoire précise, " *Les habitants de Par n'ont pas de goûts ou de préférences qui les distingueraient et forgeraient leurs personnalités*". On est alors en droit de se demander si le raisonnement de Parsifal, utilisant un raisonnement asymétrique en pensant (temporairement) pouvoir faire autre chose qu'un autre Parangon, est correct? On doit se rappeler que les raisonnements asymétriques sont utilisés dans le dilemme du prisonnier ou encore le dilemme du voyageur. Si l'on considère que l'utilisation d'un raisonnement asymétrique, bien que l'on suppose la "symétrie" des agents (ici des Parangons), est correcte, alors on doit accepter le raisonnement parangon suivant:

1. Si je suis face, soit à un Paresseux, soit un Parangon comme moi, j'ai intérêt à jouer $P_2=7/12$ (figure 1).
2. Mais, si je suis face, soit à un Paresseux, soit un Parangon qui joue $P_2=7/12$, j'ai intérêt à jouer $P_2=1$ (figure 2).
3. Mais, si je suis face, soit à un Paresseux, soit un Parangon qui joue $P_2=1$, j'ai intérêt à jouer $P_2=0$ (figure 3).
4. Mais, si je suis face, soit à un Paresseux, soit un Parangon qui joue $P_2=0$, j'ai intérêt à jouer $P_2=1$ (figure 4). Retour à l'étape 3...

Donc un Parangon se retrouve incapable de décider rationnellement d'une stratégie. La surface de probabilité de gagner de la figure 6 montre qu'il existe un minimum global en $P_2=P_2'=0$. En revanche, il existe deux maximums locaux : (1;0) et (0;1). Cependant (1;0) est bien plus favorable que (0;1). Aussi, chacun des Parangons veut jouer $P_2=1$ avec l'espoir (bien vain) que l'autre (s'il est un Parangon) jouera $P_2'=0$.

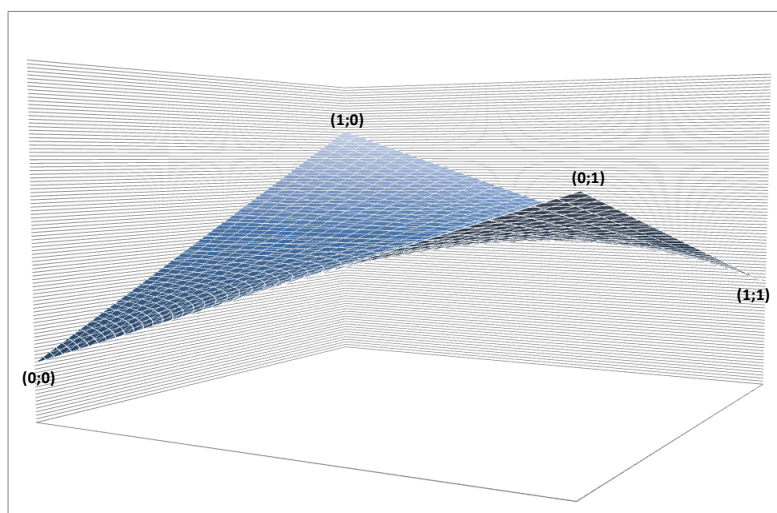


Figure 6 : surface des probabilités de gagner avec 2 Parangons asymétriques

On retrouve ici précisément le raisonnement du dilemme du prisonnier. Chacun espérant que l'autre ne dénoncera pas alors que lui-même dénoncera. On connaît la réponse de la théorie des jeux : les deux prisonniers doivent "rationnellement" dénoncer d'autre. Une solution peu efficace mais "rationnelle" au sens de la théorie des jeux. Mais ici, que doit faire Parsifal? Que doit faire un Parangon?

Discussion...

Le dilemme de Parsifal semble bien délicat. A moins... A moins que le raisonnement du Parangon ne soit pas rationnel... A moins que l'utilisation d'un raisonnement asymétrique, produisant une impossibilité de décision, soit si inopérant qu'il ne puisse être cohérent et donc soutenu.

Le raisonnement, source du dilemme, suppose une dissymétrie possible dans les décisions des Parangons. Cette dissymétrie pourrait avoir deux origines, l'une est la non-symétrie des rationalités (qui est ici équivalente au doute sur la symétrie des rationalités entre Parangons). L'autre est qu'un raisonnement conditionnel n'est pas un raisonnement situationnel (Voir Gerville-Réache 2015). En effet, le raisonnement de Parsifal "sachant qu'un Parangon jouera $P_2=7/12$, j'ai intérêt à jouer $P_2=1$ " est fondamentalement différent du raisonnement "si un Parangon joue $P_2=7/12$, j'aurais intérêt à jouer $P_2=1$ ". Le raisonnement "sachant" donne clairement un avantage à Parsifal, une dissymétrie réelle qui lui permettrait de réagir à la décision de l'autre et d'être en situation dominante.

Le dilemme est celui d'un Parangon rationnel ayant un doute sur la rationalité de l'autre. Si tous les Parangons sont identiquement rationnels et que cela est connaissance commune, il n'y a pas de dilemme et chaque Parangon jouera $P_2=7/12$. Il n'y a alors pas de dilemme entre Parangons rationnels ayant connaissance commune de leur rationalité, pas de dilemme non plus dans le dilemme du prisonnier (qui ne se dénonceraient ni l'un ni l'autre). Aussi, la théorie des jeux ne souffre pas de ces dilemmes, elle souffre de la conviction que ces dilemmes persisteraient entre agents identiquement rationnels ayant connaissance commune de leur rationalité.

Le dilemme de Parsifal pose la question des conséquences de l'hypothèse de la connaissance commune de la rationalité (CKR), de la croyance en une rationalité unique et, in fine, de la non-utilisation d'un théorème (peu connu) de Nash de 1950 qui dit que "Tout jeu symétrique admet un équilibre symétrique". Et si... deux agents rationnels ayant connaissance commune de leurs rationalités devaient théoriquement, dans tout jeu symétrique, procéder d'abord par élimination des issues non-symétriques avant d'éliminer les stratégies fortement dominées...

La question de fond est alors de comprendre dans quelle mesure l'hypothèse de la connaissance commune de la rationalité est, en elle-même, théoriquement admissible mais pratiquement inadmissible?

Ken Binmore écrit à propos du dilemme du prisonnier : "The game-theoretic equivalent of squaring the circle consists of justifying the use of a strongly dominated strategy in the one-shot Prisoner's Dilemma"... Dans le dilemme de Parsifal, l'utilisation d'un raisonnement asymétrique conduit à une impossibilité de décision. La seule issue est d'accepter :

- soit, sous l'hypothèse d'une connaissance commune de la rationalité, la symétrie comme principe premier de sélection des stratégies.
- soit l'inadmissibilité de l'hypothèse d'une rationalité unique.
- soit l'inadmissibilité de l'hypothèse d'une connaissance commune de la rationalité.
- soit l'existence d'un dilemme rationnel (voir Priest G. (2002))

L'hypothèse de connaissance commune de la rationalité nécessite que les agents soit identiquement rationnels et que cela soit connaissance commune. Cette hypothèse interdit alors tout raisonnement asymétrique dans un jeu symétrique. En effet, un raisonnement asymétrique produirait une solution temporaire théoriquement inadmissible. Aussi, soit on travaille sous l'hypothèse de connaissance commune de la rationalité et la solution dans le dilemme de Parsifal est de jouer 2 avec probabilité 7/12 et la solution est de coopérer dans le dilemme du prisonnier. Soit on travaille sans l'hypothèse de connaissance commune de la rationalité et il conviendra vraisemblablement de trahir dans le dilemme du prisonnier et de jouer 2 dans le dilemme de Parsifal.

Ken Binmore écrit également "Un argument n'est pas rationnel parce qu'il est accepté par une personne rationnelle. Au contraire, une personne est rationnelle parce qu'elle accepte uniquement des arguments rationnels."

Bibliographie

[1] Binmore, K. (1999), *Jeux et théorie des jeux*, Bruxelles : De Boeck Université.

[2] Gerville-Réache L., (2015). *Quand la probabilité conditionnelle croise la statistique*, CFIES 2015, Bordeaux, 6p.

[3] Priest G., (2002) *Rational dilemmas*, Analysis 62.1, pp. 11–16.