

**GESTION DE LA PRODUCTIVITE D'UNE FOULE**  
**PERFORMANCE ET RISQUE DE PERTES DURANT ET ENTRE LES**  
**TACHES**

<b>1. Performance durant les projets de science citoyenne : capitalisation par agrégation et capitalisation croisée.....</b>	<b>182</b>
1.1. Spécificité des projets de science citoyenne : délégation des coûts vs fiabilité du système.....	182
1.2. Améliorer la fiabilité par redondance.....	184
1.3. Répétition, perte et capitalisation.....	186
1.4. Perte durant les tâches avec exploration de l'espace : notion de capitalisation croisée.....	187
<b>2. Performance entre les tâches : capitalisation séquentielle .....</b>	<b>190</b>
2.1. Limiter les pertes entre les tâches de type élémentaire ou recette .....	190
2.2. Perte entre les tâches avec exploration de l'espace : le principe de capitalisation séquentielle.....	191
<b>3. Synthèse des situations de gestion de l'ouverture.....</b>	<b>193</b>

## **RESUME DU CHAPITRE 6**

Dans ce chapitre, nous utilisons le modèle formel de tâches développé dans le chapitre 5 pour étudier la performance de l'ouverture de ces tâches dans le cadre des projets de science citoyenne. Les projets de science citoyenne sont organisés sous une forme *seeker-solver*, où les *seekers* (les organisateurs) conçoivent et formulent la tâche et les *solvers* (les participants ou les citoyens de la science) exécutent ou résolvent la tâche. Nous montrons que l'intérêt majeur des projets de science citoyenne est l'externalisation d'une partie des coûts d'exécution des tâches aux citoyens de la science, réduisant fortement le coût global de la tâche. De plus, les tâches peuvent être parallélisées et donc diviser le temps d'exécution de tâches similaires par le nombre de participants.

En revanche, l'ouverture à la foule des tâches réduit la fiabilité du processus d'exécution de la tâche : en effet, les organisateurs n'ont pas moyen de connaître à l'avance les compétences des participants ni leur propension à fournir un effort suffisant pour résoudre la tâche. Pour pallier à cette baisse de fiabilité, les organisateurs peuvent mettre en place un système de redondance des tâches (répéter plusieurs fois la même tâche). Cette redondance des tâches peut mener cependant à des pertes importantes de productivité, pouvant parfois atteindre jusqu'à 100% de perte (par exemple dans les projets compétitifs où seule la meilleure solution est conservée).

Nous proposons de gérer ces pertes par l'introduction d'un processus de capitalisation. Dans le cas des tâches de type élémentaire ou recette, la capitalisation peut se faire par agrégation statistique, dont son fonctionnement est illustré avec le projet *Galaxy Zoo*. Dans le cas des tâches exploratoires comme la résolution de problèmes ou les tâches couplées inventives, les pertes peuvent être gérées par deux types de capitalisation : la capitalisation croisée, où les participants capitalisent durant la tâche sur ce que les autres participants ont produit. La capitalisation séquentielle : dans un premier temps, les participants explorent les espaces en soumettant plusieurs fois des solutions. A chaque solution soumise, ils améliorent leur connaissance de l'espace et peuvent capitaliser sur cette connaissance pour soumettre des solutions de meilleure qualité. Une fois la tâche terminée, il est possible de cartographier l'ensemble de l'exploration réalisée par chacun des participants. Cette cartographie permet de capitaliser sur ce qui a été produit durant la tâche et de le réutiliser dans la tâche successive.

Nous avons montré dans le chapitre 5 que les tâches scientifiques ouvertes à la foule peuvent être interprétées suivant un modèle à quatre niveaux : la tâche élémentaire, la recette, la résolution de problèmes et la tâche couplée inventive. Nous avons suggéré que la notion de performance dans ce modèle relève à la fois des questions classiques de coût-qualité-délai relatives à l'exécution des tâches ainsi qu'à l'exploration des espaces, mais intègre également d'autres critères dès lors que la tâche est répétée. Comment s'organise dans ce modèle la répartition du travail entre les différents acteurs des projets de science citoyenne ? Pour répondre à cette question, nous avons besoin de modéliser les acteurs de la tâche.

Les projets de science citoyenne font intervenir essentiellement deux types d'acteurs dans le processus, un *seeker* et un *solver* (Sieg et al., 2010). D'un côté il y a le *seeker* ou « organisateur »  $O$  dont le rôle est de concevoir, déléguer et piloter l'exécution de la tâche. C'est lui qui est en charge de gérer les critères de performance. De l'autre côté il y a les *solvers* ou « participants », qui vont exécuter la tâche. Nous définissons l'espace  $P(n) = \{P_1, P_2, \dots, P_n\}$  comme l'ensemble des participants avec  $n$  le nombre de participants. Toute personne est définie comme participant à la tâche dès lors qu'il soumet une proposition ou qu'il fait partie d'une équipe qui a soumis une proposition. Dans le contexte des projets de science citoyenne,  $P$  est un espace extensible. En effet les projets de science citoyenne ne limitent pas généralement le nombre participant et il peut toujours y avoir un nouveau participant. Le nombre  $n$  n'est donc généralement pas connu à l'avance par le *seeker*. Dans l'exemple du labyrinthe, le *seeker* est celui qui va définir le contexte (le labyrinthe), fournir les règles de déplacement (se déplacer sur les cases adjacentes et non en diagonale), définir l'objectif de la tâche (trouver la sortie), mais également définir un budget et des contraintes de réalisation. Les *solvers* soumettent des plans d'action au *seeker*, c'est-à-dire des chemins dans le labyrinthe.

Notre travail s'inscrit dans les recherches menées sur les modes de coopération entre concepteurs dans le cadre de partenariats d'exploration qui utilisent la théorie de la conception comme cadre d'analyse. Les chercheurs ont exploré différentes situations expérimentales de collaboration entre institutions dans le cas où l'incertitude est très élevée. Une forme de coopération consiste à identifier au sein de chaque partenaire des zones d'intérêts communes d'exploration : à partir d'une approche dite *matching-building* (Gillier et al., 2010), les concepteurs évoluent par étape afin de converger autour d'un concept sur lequel ils trouvent un intérêt commun. Une deuxième forme de coopération concerne les situations où différents concepteurs doivent collaborer dans des projets à très forte incertitude (inconnu marché et inconnu technologique). La stratégie proposée consiste à développer un « inconnu commun » pour gérer le risque (Kokshagina et al., 2012). Cet inconnu commun est défini comme le point de rencontre des connaissances désirables par l'ensemble des participants ou concepteurs : dès lors que cet inconnu existe, chacun d'entre eux peut générer à partir une valeur propre grâce à ses connaissances et ses expertises spécifiques. Une troisième forme de coopération consiste à faire interagir des acteurs dont les domaines d'expertises sont très éloignés. En forçant la collaboration par interactions répétées, les concepteurs peuvent entrevoir des concepts ou des idées, pourtant banales dans le domaine

initial, et qui apportent un regard neuf et désirable dans leur champ d'expertise (Salgueiredo & Hatchuel, 2016).

Dans ces exemples, l'attention est portée sur la qualité de la coopération entre acteurs, et les chercheurs ne s'intéressent pas à l'efficacité de la coopération et à la productivité. Or, de par la structure des projets de science citoyenne sous forme de seeker-solver et de challenges, la question de la performance est omniprésente : la question n'est pas tant de s'intéresser à la qualité de la coopération mais plutôt de l'impact de cette coopération sur la qualité de la production. Par ailleurs, la variété des expertises et la non connaissance *ex ante* de celles-ci par les seekers ne permettent pas de mettre en place les stratégies proposées présentées et demande de nouvelles investigations.

Dans ce chapitre, nous étudions la question de la performance dans ce modèle classique seeker-solver pour les projets de science citoyenne. Nous présentons dans la première section la gestion de la performance durant le projet. Dans la deuxième section, nous nous intéressons à la gestion des pertes entre les projets.

## **1. PERFORMANCE DURANT LES PROJETS DE SCIENCE CITOYENNE : CAPITALISATION PAR AGREGATION ET CAPITALISATION CROISEE**

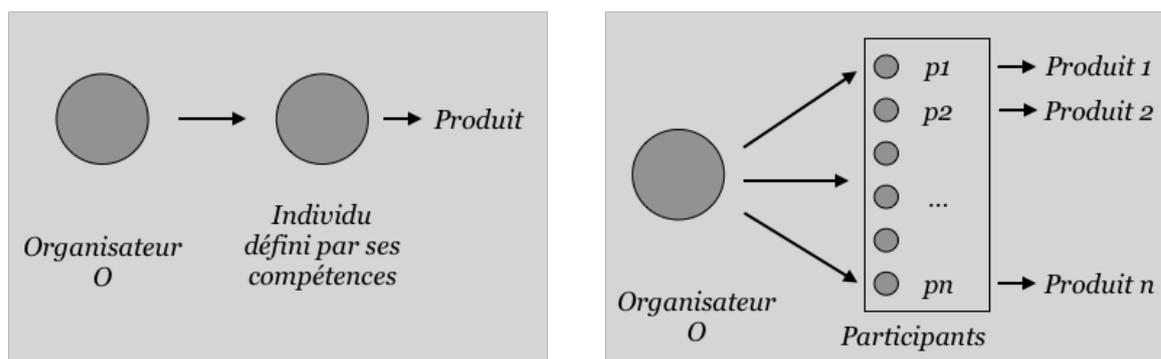
### **1.1. SPECIFICITE DES PROJETS DE SCIENCE CITOYENNE : DELEGATION DES COUTS VS FIABILITE DU SYSTEME**

Dans le modèle seeker-solver, la performance de la tâche est gérée par l'organisateur (le seeker). C'est lui qui conçoit la tâche, établit le cahier des charges à respecter dans la réalisation de cette tâche et évalue à la fin l'ensemble du processus. Nous avons vu que les critères de performance utilisés peuvent être ramenés généralement à un triptyque qualité-coût-délai. Le coût de la tâche peut être représenté comme la somme d'un coût de conception et un coût d'exécution (soit  $\text{coût} = \text{coût}_{\text{conception}} + \text{coût}_{\text{exécution}}$ ). Le coût de conception comprend l'ensemble des éléments nécessaires à la tâche et à son exécution. Ce sont notamment les ressources humaines nécessaires à la conception de la tâche, les outils associés à l'exécution de la tâche (plateforme virtuelle pour les challenges, instrument de mesure scientifique), la communication pour rendre visible la tâche et intégrer les citoyens de la science, le coût de transfert des solutions des participants vers les organisateurs. Le coût d'exécution lui comprend l'ensemble des moyens nécessaires à l'exécution de la tâche. Ce sont notamment le temps de calcul alloué à un serveur ou un ordinateur, les ressources humaines nécessaires pour exécuter la tâche ainsi que toute forme de récompense associée à l'exécution de la tâche.

Nous avons vu dans la partie 1 qu'un des avantages principaux des projets de science citoyenne est l'externalisation d'une partie des coûts d'exécution de la tâche. En effet, l'ouverture de la tâche permet au moins de déléguer les coûts relatifs aux ressources humaines nécessaires pour exécuter

la tâche. Dans le cas de Galaxy Zoo par exemple, la délégation a permis aux organisateurs d'économiser l'équivalent de 83 années de salaires pour une personne travaillant à temps plein sur le projet (Franzoni & Sauermann, 2014) - soit l'équivalent de 1.5 millions € pour une personne payée au smic. Il faut bien sûr ôter à ce chiffre les surcoûts de conception comme la création de la plateforme virtuelle, toutefois les gains en terme de coûts restent importants. D'autres projets exploitent ce gain pour réduire le temps de calcul de leurs serveurs en externalisant le calcul sur les ordinateurs individuels des citoyens de la science (voir les projets *LHC@home* et *Foldit@home*).

Un autre avantage des projets de science citoyenne est de pouvoir réduire le temps d'exécution d'un ensemble de tâches similaires. En multipliant le nombre de composants du système (les participants), l'ouverture des tâches permet d'exécuter ou de résoudre plusieurs tâches en parallèle. Le temps d'exécution de toutes les tâches est donc globalement divisé par n, n étant le nombre de participants au projet. En reprenant l'exemple de Galaxy Zoo, les participants ont codé les 900 000 images de galaxies en quelques mois, au lieu de 83 ans pour un scientifique seul.



**Figure 18. Système 1 (à gauche) : tous les composants du système sont maîtrisés, assurant sa fiabilité. Système 2 (à droite) : pas de maîtrise des caractéristiques des composants du système (les participants).**

Si la délégation de la tâche à la foule réduit les coûts et les délais en terme d'exécution de la tâche, elle réduit en même temps la maîtrise de la qualité de ce qui est produit. Contrairement à une organisation classique où la tâche est déléguée en interne à un individu dont les compétences sont reconnues, ici les organisateurs ne connaissent ni les compétences des participants, ni leur motivation à réaliser la tâche. On peut se représenter l'organisation comme un système avec les caractéristiques de la tâche à l'entrée et le produit final de la tâche à la sortie. Chaque acteur peut être représenté comme un composant de ce système dont la fiabilité a des conséquences sur la performance globale du système (**figures 17 et 18**). On passe d'un premier système 1 fiable (mais plus coûteux et plus long) où tous les composants sont maîtrisés à un système 2 moins coûteux mais moins fiable où une partie des caractéristiques des composants du système ne peut être maîtrisés.

## **1.2. AMELIORER LA FIABILITE PAR REDONDANCE**

Pour améliorer la fiabilité d'un système de production, plusieurs stratégies peuvent être mises en œuvre par les organisateurs : mise en place d'une redondance des éléments du système, amélioration de la fiabilité des sous-éléments qui constituent le système (mais potentiellement avec des coûts plus élevés), nouvelle conception du système, ou encore faire une combinaison de tout ce qui précède (E. Fyffe, W. Hines, & Kee Lee, 1968). Les projets de science citoyenne présentent toutefois des contraintes qui limitent les champs d'action pour améliorer sa fiabilité. Premièrement, il n'est pas possible de concevoir un nouveau système. Cela reviendrait à changer la foule par un autre composant du système, et donc à passer à une autre forme d'organisation. Deuxièmement, la maîtrise des composants du système est délicate à mettre en œuvre. D'un côté elle peut être gérée en augmentant les barrières à l'entrée pour participer au projet, ce qui est contraire au modèle d'ouverture promulgué par les partisans des sciences citoyennes (Franzoni & Sauermann, 2014). Une autre stratégie consisterait à cibler la communication sur le projet dans des environnements maîtrisés (des universités, des associations de spécialistes). Cette approche peut augmenter la fiabilité globale du système mais limite le nombre de participants possibles.

Les organisateurs peuvent également mettre en place une redondance dans les composants du système. La redondance consiste à disposer plusieurs composants identiques dans un système qui ont les mêmes fonctions. Elle permet d'améliorer la fiabilité globale du système de deux manières :

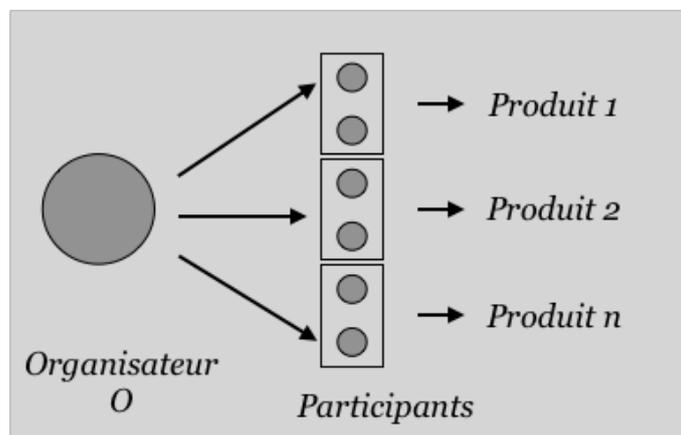
- elle réduit le risque de panne : si un des composants tombe en panne, le système peut passer par un autre composant.
- Elle augmente les performances du système. Le système peut réaliser la même action plusieurs fois par des composants différents. Un système d'évaluation peut être mis en place ensuite pour choisir le produit qui est de la meilleure qualité.

Le principe de redondance est courant dans des systèmes où les exigences en terme de sécurité sont élevées. Dans les centrales nucléaires de production électrique par exemple, la conception des systèmes de sûreté des réacteurs est réalisée conformément à des calculs statistiques pour minimiser les risques de pannes ou d'incident nucléaire. C'est le cas notamment des groupes électrogènes de secours à moteur diesel. En fonctionnement normal, le réacteur nucléaire est alimenté électriquement par des sources électriques externes. En cas de perte de ces sources, des groupes électrogènes de secours sont utilisés pour alimenter en électricité et permettre le fonctionnement des systèmes de sauvegarde qui seraient mis en œuvre en cas d'accident. Tous ces systèmes de sûreté sont doublés ou triplés, voire quadruplés tel que si l'un d'eux ne fonctionne pas, le système en réserve s'y substitue.

Bien que la redondance améliore fortement la fiabilité d'un système, celle-ci est souvent accompagnée de coûts importants. Dans les systèmes de sûreté dans les centrales nucléaires, le

seul coût partagé dans la construction des groupes électrogènes est celui de la conception du système. En revanche, la construction et l'entretien de ces systèmes sont indépendants et le coût de la redondance est environ égal au coût d'un système multiplié par le nombre de système  $coût_{redondance} \approx n \times coût_{composant}$ .

Dans les projets de science citoyenne en revanche, nous avons vu que le coût peut être en partie externalisé, donc non supporté par les organisateurs. En fait, il est possible de mettre en place un système de redondance qui s'appuie sur le fait que le nombre de participants est très grand, donc au-delà de ce qu'il est nécessaire pour réaliser la tâche. Au lieu de déléguer une tâche unique à chaque participant, les organisateurs peuvent mettre en place une redondance dans le système pour faire exécuter plusieurs fois la même tâche à des participants différents. Ainsi, le coût de la redondance n'augmente pas ou très peu : il est dépendant du coût d'exécution associé à chaque participant.



**Figure 19. Principe de redondance dans les projets de science citoyenne pour augmenter la fiabilité du système.**

Le cas du projet Galaxy Zoo (**annexe 1**) est un exemple typique d'utilisation de la redondance. Pour rappel, la plateforme Galaxy Zoo est un projet ayant regroupé plus de 250 000 volontaires qui ont aidé au codage d'images astronomiques issues de télescopes, et qui ont contribué à la découverte de nouvelles classes de galaxie. La participation est devenue rapidement virale, et sept mois après le lancement du projet environ 900 000 galaxies furent codées. A titre de comparaison, 50 millions de classifications auraient requis plus de 83 années à plein temps pour un scientifique seul. Afin de réduire la probabilité d'un codage incorrect et donc de mauvaise qualité, les galaxies furent codées plusieurs fois par différents volontaires, pour un total d'environ 50 millions de classifications, soit environ 55 codages différents par image (**tableau 7**). Une fois les galaxies codées, les organisateurs ont analysé les codages groupés des participants sur chaque image grâce à des outils statistiques et ont choisi un codage pour une image de galaxie.

### Gestion de Galaxy Zoo

<b>Tâche à exécuter</b>	Coder manuellement 900 000 images de galaxies
<b>Division en sous-tâches</b>	Coder individuellement chaque image de galaxie suivant une séquence d'actions prédéterminée
Séquence d'états pour chaque sous-tâche	Classification des galaxies prédéfinie via un arbre décisionnel: <ul style="list-style-type: none"> <li>- lisse, caractéristiques ou disque, étoile ou artefact</li> <li>- complètement rond, entre les deux, en forme de cigare,</li> <li>- ...</li> </ul>
Actions à réaliser	Evaluer visuellement les caractéristiques de la galaxie sur chaque image
<b>Gestion de la qualité de l'exécution</b>	Limiter la probabilité d'un codage incorrect
Redondance	Codage de 50 millions d'images au lieu de 900 000 uniques et évaluation par un modèle statistique

**Tableau 7. Gestion de la délégation d'une tâche de type recette : le cas de Galaxy Zoo.**

### **1.3. REPETITION, PERTE ET CAPITALISATION**

Dans l'exemple de Galaxy Zoo, la redondance est gérée par un processus de **capitalisation par agrégation**. Ce modèle de capitalisation est celui que nous avons présenté dans le chapitre précédent la notion de capitalisation appliqué durant la tâche. En fait, nous pouvons rapprocher de manière plus générale le principe de redondance à la notion de tâche répétée. Créer une redondance dans l'exécution d'une tâche revient exactement à répéter plusieurs fois la même tâche au sein d'un système. Les questions de performance que nous avons présentées dans le cas où la tâche est répétée s'appliquent dès lors qu'il y a redondance dans le système de production de la tâche.



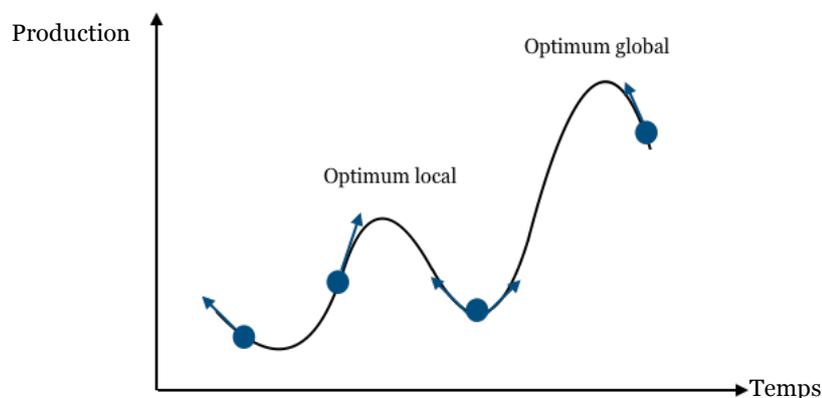
Nous avons vu que la répétition des tâches peut mener à des variations de performance entre les tâches et durant la tâche, comme des problèmes de pertes. Dans le cas de la délégation à la foule, ces pertes peuvent être très importantes. En effet, la fiabilité du système dans les tâches ouvertes à la foule repose souvent sur une approche statistique (concordance statistique sur des tâches répétées dans le cas de la recette, augmentation de la probabilité de trouver une bonne réponse pour la résolution de problèmes, multiplication des explorations dans le cas des tâches couplées inventives) et une grande partie de ce qui est produit par les participants n'est pas utilisé. Dans le cadre de la résolution de problème par exemple, chaque participant travaille individuellement et seule la meilleure solution est conservée. Toutes les autres productions peuvent être vues comme un ensemble de mauvais chemins qui n'ont pas servi à l'élaboration de la solution finale : le taux

de perte est proche de 100% ( $\frac{n-1}{n}$  solutions non utilisées). D'un point de vue économique, une telle perte est un non sens pour la productivité.

#### **1.4. PERTE DURANT LES TACHES AVEC EXPLORATION DE L'ESPACE : NOTION DE CAPITALISATION CROISEE**

##### **1.4.1. Redondance et perte dans la résolution de problèmes**

Dans le cas de la résolution de problèmes, le mode d'organisation par redondance permet de multiplier les positions de départ dans l'exploration des solutions, et donc d'augmenter la probabilité d'être proche de la solution optimale (Afuah & Tucci, 2012; King & Lakhani, 2013). Chaque participant va explorer une partie de l'espace des plans d'action créant une multitude de recherches locales (Afuah & Tucci, 2012). L'organisation de la tâche est donc basée sur *l'augmentation de la diversité des solutions proposées*. Ce type d'organisation a fait l'objet de plusieurs études dans le cas de la résolution de problèmes et plusieurs chercheurs ont suggéré son intérêt pour obtenir des solutions de meilleure qualité (voir e.g. Amatriain, 2012; Boudreau, Lacetera, & Lakhani, 2011; Brabham, 2008; Cavallo, Street, York, & Haven, 2012; Poetz & Schreier, 2012). En plus de fournir des solutions plus performantes qu'une exploration individuelle (Poetz & Schreier, 2012), ces projets sont peu coûteux car les organisateurs ne rémunèrent généralement que le vainqueur.



**Figure 20. Multiplication des points d'entrée et des chemins d'expérimentation dans l'espace des plans d'action.**

En revanche, l'organisation compétitive pour la résolution de problèmes présente l'inconvénient majeur de générer une perte importante de productivité. En effet, seule la séquence d'actions du vainqueur est exploitée, tandis que les autres séquences ne sont généralement jamais divulguées aux organisateurs. En terme de rentabilité, le ratio de perte de production est proche de 100% lorsque le nombre de participants est grand. Pour limiter cette perte, des économistes et des gestionnaires suggèrent de nouveaux modèles d'organisation incluant une coopération entre les participants (Afuah, 2018; Benkert & Letina, 2016; Boudreau & Lakhani, 2015). Au lieu que

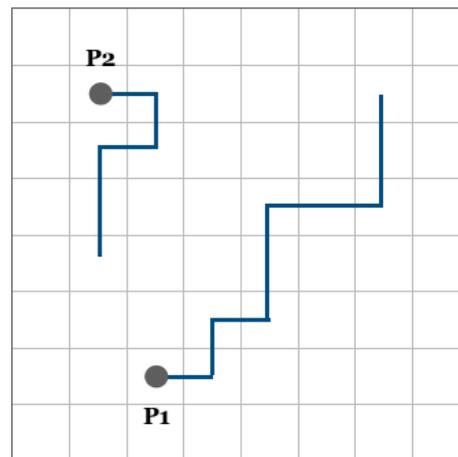
chaque participant travaille individuellement, les participants peuvent avoir accès à tout ou partie de ce qui a été produit par les autres. Ce système est compatible avec les projets de science citoyenne, où les organisateurs ouvrent les résultats intermédiaires durant le processus. Chaque citoyen de la science a accès aux séquences d'action que les autres participants ont produit, et peut les utiliser sans contrainte particulière. Contrairement au modèle compétitif, l'approche collaborative de la résolution de problèmes diminue le risque de perte d'une « bonne idée » qui peut être réutilisée par les autres participants. Ce modèle d'organisation a été montré comme donnant des solutions de meilleure qualité que le processus compétitif dans les mêmes circonstances (Boudreau & Lakhani, 2015). Or, si l'ouverture des résultats intermédiaires augmente les performances du processus les mécanismes d'interactions entre les participants sont peu clairs. En effet, l'étude réalisée par Boudreau et Lakhani montre que si la meilleure solution est de meilleure qualité, il y a moins de diversité dans les solutions, contredisant le modèle compétitif. Alors que les modèles de gestion de la résolution de problèmes par la foule se basent généralement sur un modèle input/output où les organisateurs n'interagissent pas durant le processus, l'ouverture à la collaboration force à repenser ce paradigme et à concevoir un modèle de gestion qui prenne en compte les interactions entre les participants durant le processus.

#### ***1.4.2. Perte durant les tâches couplées inventives***

Les études de performance menées sur l'ouverture des tâches de type résolution ne s'appliquent pas dans le cas des tâches couplées inventives. En effet, ces études suggèrent que la performance de l'ouverture à la foule nécessite l'existence d'un problème bien formulé, c'est-à-dire qu'il existe un état désirable formulé. Cependant, nous pouvons étendre les notions de perte à l'exploration de l'espace des états et celui des plans d'action. Dans le cas d'une tâche couplée inventive compétitive, chaque participant explore de manière individuelle les deux espaces. Cette exploration permet d'augmenter la diversité des zones des espaces explorées. En effet, nous avons vu dans le chapitre précédent que le nombre d'hypothèses possibles croît exponentiellement avec le nombre de variables à combiner. Dans le cas où la tâche est déléguée à une foule, la grande taille de l'espace des hypothèses diminue la probabilité que deux participants formulent la même hypothèse. En revanche plus il y a de participants à explorer les espaces, plus il y a de chances que certaines zones explorées des espaces aient des parties en commun. Par exemple, deux participants travaillant sur des hypothèses différentes mais dont les relations ainsi qu'une partie des objets sont identiques. Or, cette zone commune pourrait être délaissée par un des participants qui n'en voit pas l'usage tandis que l'autre y aurait un intérêt dans son exploration. L'exploration menée par le premier participant serait dans ce cas une pure perte. Ces risques de perte durant la tâche sont moins importantes que pour la résolution de problèmes car ils dépendent de multiples facteurs : existence d'une zone commune d'exploration, variation dans les intérêts sur cette zone commune. C'est pourquoi notre étude se concentrera essentiellement sur les tâches de type résolution de problèmes.

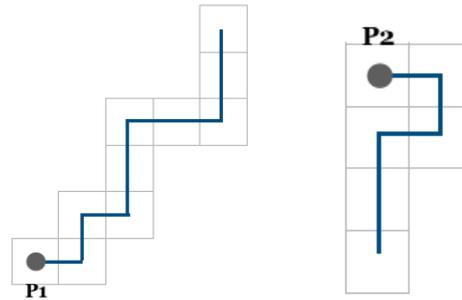
### 1.4.3. Limiter les pertes par capitalisation croisée

Alors que la résolution compétitive de problèmes est basée sur la diversité des solutions proposées, l'organisation collaborative cherche à diminuer les pertes en améliorant la *capitalisation* sur ce qui est produit par les participants. Pour améliorer la performance globale du processus de résolution de problèmes, la gestion de la capitalisation est donc un élément central que la littérature n'a que peu étudié dans le cadre d'une résolution collective. En effet, la réutilisation entre les participants, que nous définissons comme **capitalisation croisée**, permet notamment de réduire le risque que plusieurs participants explorent plusieurs fois la même zone de l'espace.



**Figure 21. Illustration des parcours de deux participants dans le labyrinthe.**

Nous pouvons illustrer la capitalisation croisée au travers de l'exemple du labyrinthe. Supposons deux participants  $P1$  et  $P2$  qui soumettent chacun une séquence d'action  $A1$  et  $A2$  respectivement. Supposons également que les participants n'ont pas trouvé de séquence qui mène à la sortie et doivent proposer une nouvelle solution. Dans le cas où la résolution est compétitive, chaque participant ne sait pas ce que l'autre a soumis. Chacun ne connaît donc que son état initial, son état final et les états intermédiaires de la séquence qu'il a soumise. La mémoire des états possibles dans le labyrinthe est individuelle. Il y a donc un risque que le participant  $P1$  soumette une nouvelle séquence dont certains états se confondent avec la séquence d'action  $A2$ . Dans notre exemple, il a 6 chances sur 64 que cela se produise. De la même manière,  $P2$  a 10 chances sur 64 de proposer un état qui aura déjà été exploré par  $P1$ . Dans le cas où la résolution est collaborative, chaque participant connaît la séquence d'action soumise par l'autre et voit toutes les zones de l'espace qui ont été explorées. La probabilité de proposer un état déjà exploré tombe à 0. De plus les deux participants connaissent chacun 16 états sur 64 de l'espace et donc ont moins d'états à tester.



**Figure 22. Dans la résolution compétitive, chaque participant ne peut capitaliser que sur sa séquence d'action.**

Quel comportement est adopté par les participants lors de la capitalisation croisée ? Est-ce qu'il réutilisent une partie des séquences d'actions soumises par l'autre participant ou bien préfèrent-ils explorer des zones de l'espace encore inconnues ? Nous allons étudier cette question dans le chapitre suivant en analysant le cas du RAMP, une plateforme de data challenge. Nous montrerons que les participants ont un comportement globalement identique qui crée des plateaux de fixation autour de certaines zones de l'espace. Nous montrerons également que la capitalisation croisée peut être améliorée si les organisateurs mettent en place des challenges compétitifs puis collaboratifs.

## ***2. PERFORMANCE ENTRE LES TACHES : CAPITALISATION SEQUENTIELLE***

### ***2.1. LIMITER LES PERTES ENTRE LES TACHES DE TYPE ELEMENTAIRE OU RECETTE***

Nous avons vu les problématiques de perte lorsque plusieurs tâches identiques sont réalisées en même temps. La redondance peut également être organisée de manière séquentielle, c'est-à-dire sans recoupement entre les tâches identiques. Dans le cas où ce sont les mêmes acteurs qui participent aux projets (du moins en partie), la transmission d'information et la gestion des pertes se fait en partie de manière tacite par des échanges entre les participants (Brady & Davies, 2004). Cependant, dès lors que les tâches sont déléguées à une foule, le risque de pertes est décuplé. En effet, les participants restent rarement sur plus d'un projet, limitant la transmission tacite d'information. Comment alors gérer cette perte dans les projets ?

Dans les tâches de type élémentaire ou recette, la partie principale de l'apprentissage concerne ce que les organisateurs peuvent apprendre de l'utilisation de citoyens dans le processus, des compétences des participants et des types d'actions qu'ils peuvent leur demander d'exécuter (Sauermaann & Franzoni, 2014). Le retour d'expérience fait par les organisateurs permet de définir les contours de la méthode utilisée et de mieux cibler pour les prochains projets ce qu'il est possible de demander à la foule et sous quelle forme la tâche doit être présentée. A titre

d'exemple, le premier projet Galaxy Zoo a permis de démontrer aux organisateurs l'utilité de la méthode pour la production de catalogues à grande échelle et de découvertes fortuites d'objets individuels (Thomas et al., 2013). Depuis lors, cette méthode a été étendue au-delà des morphologies de galaxies pour inclure l'identification de la supernova (Smith et al., 2011), la découverte d'exoplanètes (Fischer et al., 2012) et un recensement des bulles associées à la formation d'étoiles dans la Voie Lactée (Simpson et al., 2012), ainsi que divers problèmes liés aux «données volumineuses» en dehors de l'astronomie avec la plateforme Zooniverse. Alors que le projet original Galaxy Zoo identifiait les galaxies comme étant des types précoces, des types tardifs ou des fusions, le projet Galaxy Zoo 2 s'est attelé à mesurer des caractéristiques morphologiques plus fines comme les barres, les renflements et les formes des disques à bords, ainsi que la quantification des forces relatives des renflements galactiques et des bras spiraux.

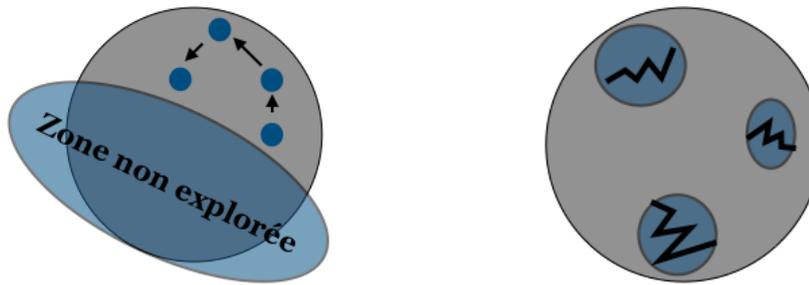
## ***2.2. PERTE ENTRE LES TACHES AVEC EXPLORATION DE L'ESPACE : LE PRINCIPE DE CAPITALISATION SEQUENTIELLE***

La transmission d'informations entre les projets est plus complexe dans le cadre de tâches incluant l'exploration d'un ou plusieurs espaces. En effet la recherche d'une solution est souvent sujette à de « mauvais chemins », de « l'expérimentation », de la « sérendipité » et de « l'incertitude » (Abernathy & Rosenbloom, 1969; Boudreau, Lakhani, & Lacetera, 2008; Loch, Terwiesch, & Thomke, 2001). La découverte d'une solution à un problème se confronte alors à la création de connaissances qui se fait souvent en tant que sous-produit involontaire de l'activité de projet (DeFillippi, Jones, & Arthur, 2001).

Supposons par exemple que la résolution d'un problème soit donnée successivement à différents participants ou groupes de participants. Durant la première résolution, chaque participant va explorer une partie de l'espace en proposant un ensemble de séquences d'action pour résoudre le problème. Nous pouvons représenter cette exploration par une succession de petits chemins dans l'espace (**figure 23**). Cette exploration va former une zone de l'espace qui aura été explorée par le participant. Si plusieurs participants explorent l'espace en même temps, nous pouvons représenter les explorations individuelles avec une cartographie de l'espace. Cela permet de distinguer les zones de l'espace explorées des zones de l'espace non explorées, ainsi que des distances possibles entre les zones de l'espace explorées (**figure 24**). Cette cartographie fournit des informations pour les tâches successivement répétées : elle représente la mémoire de ce qui a déjà été explorée dans l'espace. Sans cette mémoire, les participants des tâches suivantes ne pourraient pas s'appuyer sur les explorations déjà faites, et donc recommenceraient l'exploration comme si rien n'avait été fait auparavant. Or, cette cartographie permettrait de savoir si certaines zones de l'espace ont été très explorées tandis que d'autres ont été négligées, et potentiellement réduire les coûts d'explorations des participants.

L'existence de deux espaces à explorer dans les tâches couplées inventives augmente d'autant plus les choix dans l'exploration et donc le risque de partir sur de mauvais chemins. Nous avons

montré dans le chapitre 5 par exemple que la génération d'hypothèses basée sur les données est associée à un risque très fort de générer des hypothèses sans intérêt scientifique.



**Figures 23 et 24. A gauche : exploration individuelle de l'espace. A droite : cartographie des explorations individuelles.**

Réduire les pertes en utilisant la production des tâches précédentes fait apparaître un processus de **capitalisation séquentielle**. Dans un premier temps, les participants explorent les espaces en soumettant plusieurs fois des solutions. A chaque solution soumise, ils améliorent leur connaissance de l'espace et peuvent capitaliser sur cette connaissance pour soumettre des solutions de meilleure qualité. Une fois la tâche terminée, il est possible de cartographier l'ensemble de l'exploration réalisée par chacun des participants. Cette cartographie permet de capitaliser sur ce qui a été produit durant la tâche et de le réutiliser dans la tâche successive.

Ce principe de capitalisation séquentielle permet de gérer la production entre les tâches et de réduire les pertes de production en utilisant successivement ce qui est produit. Cependant, si le principe de capitaliser sur ce qui a été produit semble assez intuitif dans un processus de performance, les moyens pour gérer cette capitalisation séquentielle doivent être éclaircis. Dans le cas d'une tâche de type résolution de problèmes, la production durant la tâche peut être résumée à la séquence d'action qui a le meilleur score à la fonction *test()*. Celle-ci peut par exemple servir de base pour la répétition successive de la tâche. Ce processus de capitalisation séquentielle est moins clair dans le cas d'une tâche couplée inventive. Qui est en charge de la capitalisation dans le processus de capitalisation séquentielle ? Quels sont les moyens à mettre en œuvre pour gérer cette capitalisation ? Nous tenterons de répondre à ces questions au travers de notre analyse du cas Epidemium dans les chapitres 8 et 9.

### 3. SYNTHÈSE DES SITUATIONS DE GESTION DE L'OUVERTURE

Le modèle que nous avons construit dans les chapitres 5 et 6 propose de distinguer quatre types de tâches déléguées par l'ouverture du processus scientifique à la foule : tâche de type élémentaire ou recette, résolution de problème, tâche couplée inventive. Chacune de ces tâches est associée à des activités différentes (exécution d'une séquence d'action, exploration d'un ou deux espaces) et demande de mettre en place des moyens de gestion différents. Nous avons montré dans ce chapitre comment la capitalisation a été gérée pour des tâches de type élémentaire ou recette au travers de l'exemple de Galaxy Zoo. Par la suite, nous allons tenter de montrer comment cette capitalisation est gérée dans les projets de science citoyenne sur des tâches de type résolution de problèmes et tâche couplée inventive.

	<b>Tâche élémentaire ou recette</b>	<b>Résolution de problèmes</b>	<b>Tâche couplée inventive</b>
<b>Type d'espace dans le modèle</b>	Un espace constitué de l'ensemble des plans d'action	Un espace constitué de l'ensemble des plans d'action	Deux espaces : - Un espace constitué de l'ensemble des plans d'action - Un espace constitué des hypothèses
<b>Activité</b>	Exécution d'une action ou d'un plan d'action prédéterminé (état du monde espéré entièrement connu)	Exploration de l'espace des plans d'action grâce guidée par l'existence d'une fonction test et d'un état final désiré	Exploration des deux espaces suivant le libre choix de l'explorateur pour construire le quoi et le comment des hypothèses
<b>Collaboration</b>	Organisation hiérarchique seeker-solver, division du travail entre les participants	Organisation hiérarchique seeker-solver, coordination par réutilisation de la production durant la tâche	Organisation hiérarchique seeker-solver,
<b>Risque de perte dans la répétition des tâches</b>	Distinguer les bonnes exécutions des mauvaises, ne pas exploiter au maximum la foule	Perte des « bonnes idées » dans l'exploration de l'espace des plans d'action durant la tâche (risque élevé), risque de recommencer l'exploration de zéro (perte entre les tâches)	Zones communes d'exploration dans les espaces (risque faible), risque élevé de produire des éléments non intéressants
<b>Capitalisation durant la tâche</b>	Capitalisation par agrégation	<b>Capitalisation croisée</b>	<b>Capitalisation croisée</b>
<b>Capitalisation entre les tâches</b>	Retour d'expérience	<b>Capitalisation séquentielle</b>	<b>Capitalisation séquentielle</b>
<b>Organisation de la capitalisation</b>	Production par les participants, Capitalisation par les organisateurs	Production et capitalisation croisée par les participants Capitalisation séquentielle gérée par les organisateurs	Capitalisation séquentielle gérée par les organisateurs ?

**Tableau 8. Synthèse des tâches au sein du processus scientifique et des types de capitalisation.**

