Les informations exigées par la législation REACH : Analyse du partage des coûts*

Sylvain Béal**, Marc Deschamps†, Joël-Thomas Ravix‡, Olivier Sautel§

7 août 2009

Résumé

Cet article s'intéresse au partage des coûts à l'intérieur de la législation REACH. Les membres d'un forum d'échange d'information ont l'obligation de partager les données qu'ils possèdent. Ce partage de l'information conduit à la mise en place d'un mécanisme de compensation : les membres non propriétaires d'une information compensent financièrement les membres propriétaires. Nous nous intéressons au choix d'un mécanisme de compensation équitable pour l'ensemble des participants. Pour cela, nous modélisons la procédure de partage des coûts de production des informations par un jeu coopératif à utilité transférable. Plus précisément, nous reprenons le formalisme des jeux de partage de données récemment proposé par Dehez et Tellone [2008]. Cinq mécanismes de compensation sont proposés et caractérisé par une collection d'axiomes. Une discussion des axiomes est proposée.

Mots clés : Législation REACH, partage de données, mécanisme de compensation, partage égalitaire, valeur de Shapley, coût marginal.
Classification JEL : C71, D71, K32, L65.

The information exchanged under REACH: A cost-sharing analysis

Abstract

This article addresses the question of cost-sharing mechanisms under REACH. The members of an information exchange forum have to share their data. This exchange requires the adoption of a compensation scheme: the owner of a data receives a monetary compensation in return for the access to his/her data. We are interested in the choice of a compensation scheme that would be fair for all participants. In order to answer this question, the exchange of information is modeled as a cooperative game with transferable utility. More specifically, we use the class of data games introduced by Dehez and Tellone [2008]. Five compensation schemes are considered and characterized by a collection of axioms. A discussion of the axioms is proposed.

Keywords: REACH regulation, data sharing, compensation scheme, equal division, Shapley value, marginal cost.
JEL classification: C71, D71, K32, L65.

*Nous remercions Pierre Bernhard, Amandine Ghintran, Gabrielle Moineville, Céline Savard-Chambard, Philippe Solal et Marta Stryszowska pour leurs commentaires et suggestions. Le premier auteur a également bénéficié du soutien financier de l'Agence Nationale de la Recherche, dans le cadre du programme "Models of Influence and Network Theory".

**Université de Saint-Etienne, CREUSET, CNRS UMR 5824, 6 rue basses des rives, 42023 Saint-Etienne. E-mail : `sylvain.beal@univ-st-etienne.fr`

†Université de Nice Sophia-Antipolis, GREDEG, CNRS UMR 6227, 250 rue Albert Einstein, 06560 Valbonne. E-mail : `marc.deschamps@gredeg.cnrs.fr`

‡Université de Nice Sophia-Antipolis, GREDEG, CNRS UMR 6227 et OFCE-DRIC, 250 rue Albert Einstein, 06560 Valbonne. E-mail : `joel.ravix@gredeg.cnrs.fr`

§Microeconomix et OFCE-DRIC, 5 rue du 4 septembre, 75002 Paris. E-mail : `olivier.sautel@microeconomix.com`

1 Introduction

Les produits chimiques sont explicitement ou implicitement présents dans un très grand nombre d'activités industrielles ou de services et, à ce titre, forment le quotidien de très nombreuses entreprises ainsi que des consommateurs. La Commission européenne estime que plus de cent mille substances chimiques sont actuellement utilisées en Europe et il semble pratiquement impossible de trouver des marchés qui n'utilisent pas, peu ou prou, des produits chimiques en tant que tels, dans des préparations, ou contenus dans des articles. Dès lors, la gestion sanitaire et environnementale de ces produits est un sujet crucial pour les pouvoirs publics, notamment du fait des préoccupations exprimées par les associations de consommateurs, les organisations non gouvernementales, les partis politiques et, de façon plus large, les citoyens européens.

Dans le but principal de répondre à ces attentes, les institutions européennes ont décidé de mettre en œuvre un nouveau cadre réglementaire harmonisé : la législation européenne REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals)[1]. L'objectif poursuivi est parfaitement exposé au premier paragraphe de l'article premier du Règlement CE n°1907/2006 du Parlement européen et du Conseil, d'après lequel : “Le présent règlement vise à assurer un niveau élevé de protection de la santé humaine et de l'environnement, y compris la promotion de méthodes alternatives pour l'évaluation des dangers liés aux substances, ainsi que la libre circulation des substances dans le marché intérieur tout en améliorant la compétitivité et l'innovation”[2]. Cette nouvelle législation vise ainsi conjointement à améliorer la connaissance des usages et des propriétés des substances chimiques fabriquées ou importées dans l'Union européenne, à assurer la maîtrise des risques qui leur sont liés, et, le cas échéant, à restreindre voire à interdire leur utilisation. Selon les institutions communautaires, environ trente mille substances chimiques sont concernées par cette législation.

A cette fin, le Règlement entré en vigueur le 1er juin 2007 pose un principe qui, par rapport au cadre réglementaire précédent, inverse la charge de la preuve de l'innocuité des produits chimiques. Désormais, ce sont les fabricants, les importateurs et les utilisateurs en aval qui doivent “veiller à fabriquer, mettre sur le marché ou utiliser des substances qui n'ont pas d'effets nocifs pour la santé humaine ou l'environnement. Ces dispositions reposent sur le principe de précaution” (art 1, §3). L'effectivité de ce principe est assurée par le fait qu'en l'absence de respect de la législation, les substances, préparations ou articles ne peuvent être commercialisés sur le territoire de l'Union européenne. Les pays membres sont responsables des contrôles ainsi que de la définition et de la mise en œuvre des amendes, sanctions administratives et pénales[3].

[1]Historiquement, la législation REACH trouve son origine dans la réunion du Conseil des ministres de l'environnement de l'Union européenne en 1998, le Livre Blanc de la Commission européenne de 2001, la proposition de Règlement votée le 17 novembre 2005 par le Parlement européen et son adoption en deuxième lecture le 13 décembre 2006. Cette nouvelle législation remplace plus de quarante Directives antérieures. Les études d'impact réalisées par la Commission européenne avant l'adoption du Règlement ont estimé qu'à terme il y aurait une diminution de deux mille à quatre mille décès pour cause de cancer par an, ainsi qu'une réduction des dépenses de santé publique qui pourrait aller jusqu'à cinquante milliards d'euros sur trente ans.

[2]L'amélioration de la compétitivité et de l'innovation réside dans le fait que les institutions européennes considèrent qu'en encourageant la substitution des substances les plus dangereuses, REACH favorise la recherche et le développement de nouvelles substances et donc la création de nouveaux emplois.

[3]Voir l'ordonnance 2009–229 du 29 février 2009 dans le cas de la France.

De façon concrète, le Règlement crée une Agence européenne des produits chimiques (ECHA) chargée de mettre en œuvre la législation REACH reposant sur quatre procédures principales : l'enregistrement, l'évaluation, l'autorisation, et la restriction. La nouvelle procédure d'enregistrement constitue l'enjeu principal et la plus grande innovation réglementaire : elle concerne, sauf exemptions expressément prévues, toutes les substances chimiques, existantes ou nouvelles, produites ou importées à plus d'une tonne par an. Cette obligation est extrêmement forte puisqu'elle contraint les firmes, lors de l'enregistrement, à fournir toutes les données existantes portant sur les propriétés des substances qu'elles utilisent. Cette procédure prévoit que sans données, et donc sans enregistrement, les entreprises ne peuvent utiliser les substances : "Sans données, pas de marché" comme le résume l'ECHA.

Face à l'immensité de la tâche à accomplir et en particulier à la complexité logistique d'une telle réforme, une phase de pré-enregistrement des substances existantes (du 1er juin 2008 au 1er décembre 2008) a été introduite. L'avantage de cette dernière est d'ouvrir l'accès à un régime transitoire permettant de reporter la date de l'enregistrement des substances chimiques (à 2010, 2013, ou 2018), en fonction de leur dangerosité et du tonnage utilisé. Du point de vue des entreprises concernées, le pré-enregistrement nécessitait : la réalisation d'un inventaire des substances utilisées, l'identification de ces substances au regard du guide technique d'identification, l'évaluation des quantités utilisées, et la détermination du statut de l'entreprise vis-à-vis de chacune de ces substances (producteur, importateur ou utilisateur aval). Il faut noter que dans l'éventualité où une entreprise n'aurait pas pré-enregistré et souhaiterait continuer à produire, elle doit immédiatement enregistrer sa substance, ce qui conduit inévitablement à stopper sa production ou sa commercialisation, au minimum durant trois semaines, le temps que l'ECHA contrôle le dossier et demande éventuellement des compléments d'informations.

Afin de fournir les données nécessaires à l'enregistrement ainsi que d'éviter la réplication des tests, notamment sur les animaux vertébrés, les instances communautaires ont prévu que dès le pré-enregistrement d'une substance une entreprise est inscrite d'office dans un Forum d'Echange d'Information sur les Substances (pré-SIEF) grâce au logiciel REACH-IT. La participation à un SIEF étant obligatoire pour toute entreprise souhaitant enregistrer une substance, celle-ci n'a le choix que du SIEF qui lui paraît le plus approprié. L'objectif d'un SIEF dans cette législation est double : il s'agit à la fois de faciliter l'échange des informations physico-chimiques, toxicologiques et écotoxicologiques afin d'éviter la duplication des tests, mais aussi de se mettre d'accord sur la classification et l'étiquetage lorsqu'il existe des différences entre les déclarants dans la proposition de classification et d'étiquetage de la substance. Au sein du SIEF, les entreprises décident librement des modalités de communication et d'organisation ainsi que de la forme juridique de celui-ci. Le Règlement impose que les membres d'un SIEF fournissent à leurs pairs les études existantes, répondent à leurs demandes d'informations et œuvrent collectivement pour identifier et réaliser des études complémentaires si cela est nécessaire. L'ensemble de ce travail au sein du SIEF vise à permettre une soumission conjointe pour la substance[4],

[4] Le Règlement prévoit, sous certaines conditions, qu'une entreprise puisse s'exclure volontairement de l'obligation de participer à la soumission conjointe. Il indique que cela peut notamment être le cas si une soumission conjointe entraîne des coûts trop élevés pour l'entreprise ou si elle l'oblige à dévoiler des informations confidentielles. Pour autant, une entreprise qui choisirait de s'exclure volontairement de la soumission conjointe reste membre d'un SIEF et doit donc continuer à partager ses données.

grâce à un déclarant principal, de façon à ce que chaque entreprise déclarante puisse s'y référer dans son dossier d'enregistrement et qu'ainsi elle n'ait plus qu'à fournir les données sur les spécificités de sa production ou de ses usages. La date d'enregistrement correspond à la date de soumission du dossier complet par le déclarant principal. Les SIEF sont opérationnels du 1er janvier 2009 jusqu'au 1er juin 2018, date de fin de la période transitoire la plus longue.

La question du partage des données au sein d'un SIEF est donc une question cruciale au moins jusqu'au 1er juin 2018. C'est la raison pour laquelle l'ECHA a proposé dès septembre 2007, à titre indicatif, un "Guide sur le partage des données", laissant aux membres de chaque SIEF la liberté de choisir le système qu'ils préfèrent tant qu'il reste "équitable, transparent et non discriminatoire".

L'objet de notre article est de montrer, notamment à la suite de Dehez et Tellone [2008], comment la théorie des jeux coopératifs permet de fournir un cadre d'analyse de la question du partage des données au sein d'un SIEF. Cette question implique de s'interroger sur la façon dont chaque donnée sera valorisée ainsi que sur la compensation permettant le transfert. Compte tenu de la très grande diversité de cas empiriquement envisageables et du fait qu'en ce domaine nous considérons que seul le "sur-mesure" permet de fournir une réponse adéquate aux spécificités des données et aux souhaits des membres d'un SIEF, nous avons retenu trois hypothèses simplificatrices. La première consiste à considérer que tous les membres d'un SIEF, qu'ils possèdent ou non des données, sont des entreprises déclarantes. La deuxième hypothèse pose que tous les membres d'un SIEF ont nécessairement besoin de toute l'information disponible. Enfin, notre troisième hypothèse porte sur le fait que toutes les informations ont la même qualité. Ces trois hypothèses simplificatrices ne nous semblent pas modifier l'aspect qualitatif de nos résultats et peuvent, dans certains cas, appréhender les aspects saillants de certains SIEF (ou de certains consortiums[5]). Nous revenons sur ces hypothèses dans la section 8.

Le formalisme des jeux de partage de données proposé par Dehez et Tellone [2008] peut être décrit de la manière suivante. Un SIEF est constitué d'un ensemble d'agents. Chaque agent possède certaines données, sachant qu'une donnée peut appartenir à plusieurs agents et qu'un agent peut ne posséder aucune donnée. Par ailleurs, chaque donnée a un coût de reproduction correspondant à l'investissement que devrait réaliser un agent afin de reproduire la donnée. Il est donc possible d'associer à chaque coalition d'agents du SIEF une quantité monétaire mesurant le coût total de reproduction des données non possédées par les membres de la coalition. La théorie des jeux coopératifs se prête particulièrement bien à l'étude du partage de l'information au sein des SIEF dans la mesure où chaque utilisateur d'une substance a l'obligation de participer au SIEF (voir le chapitre 5 de Moulin [1981] pour cette interprétation). A ce titre, le pouvoir de décision des participants est remis, indivisible, entre les mains de la collectivité chargée d'arbitrer souverainement les conflits d'opinion de ses membres. La quantité monétaire associée à l'ensemble de tous les membres d'un SIEF est nulle puisque ces agents n'ont aucune donnée à reproduire. Ceci signifie que le jeu de partage de données est une situation de pure compensation dans le sens suivant : une règle de partage pour ce jeu est un mécanisme de compensation

[5]Le Règlement autorise en effet les membres d'un même SIEF, ou de plusieurs SIEF, à avoir une coopération renforcée au sein d'un consortium, structure juridique contractuelle qui dans de nombreux cas a été mise en œuvre par les industriels avant ou durant la phase de pré-enregistrement (voir Vincent et Garderes [2008] sur ce point).

sur lequel les agents non propriétaires de données offrent une compensation aux agents propriétaires en échange de l'accès aux données.

Cet article est consacré à l'étude de cinq de ces mécanismes inspirés de la théorie des jeux coopératifs. Les mécanismes proposés couvrent un large éventail de principes de justice distributive : certains reposent sur des principes égalitaristes, d'autres sur des principes marginalistes. Notre démarche est axiomatique : on liste une collection de propriétés, jugées désirables par la collectivité, que doit satisfaire un mécanisme de compensation, puis on caractérise, si possible, l'ensemble des mécanismes de compensation satisfaisant ces propriétés. L'objectif est de définir un système d'axiomes conduisant à un unique mécanisme de compensation. La démarche axiomatique est utilisée dans de nombreux contextes économiques (voir Thomson [2001]). Cet article se veut opérationnel au sens où un soin tout particulier est accordé à la présentation des axiomes afin qu'ils puissent guider le praticien dans le choix d'un mécanisme de compensation.

Le reste de l'article est organisé en sept sections. La section 2 consiste à identifier certaines des propriétés désirables pour les membres d'un SIEF, que doit satisfaire un mécanisme de compensation. La section 3 présente le formalisme des jeux de partage de données. La section 4 utilise ce formalisme afin de définir précisément les propriétés envisagées dans la section 2. Cinq mécanismes de compensation possibles sont introduits dans la section 5. Les quatre premiers sont classiques dans la littérature portant sur ce type de jeux, le dernier est un nouveau mécanisme que nous avons appelé "mécanisme de compensation intégrale". La section 6 est consacrée à la caractérisation de ces cinq mécanismes de compensation. Un exemple numérique illustrant les résultats obtenus est proposé dans la section 7. La section 8 conclut par une ultime comparaison de nos cinq mécanismes de compensation et discute le mécanisme de compensation Klimisch figurant explicitement dans le Guide de l'ECHA.

2 Quelques propriétés souhaitables

Une négociation entre des agents peut se terminer par de nombreuses issues très différentes. Il est donc difficile de définir une unique propriété désirable pour les négociants, que devrait vérifier le résultat de chaque négociation. Plusieurs principes de justice distributive peuvent apparaître désirables pour la collectivité. Une négociation entre des membres d'un SIEF à propos des coûts des informations échangées n'échappera pas à ce principe. Dans cette section, nous listerons donc une collection de propriétés permettant de décrire des mécanismes de compensation résultant d'une telle négociation. Certaines de ces propriétés seront compatibles, d'autres seront inconciliables. De plus, certaines propriétés sembleront incontournables au lecteur, d'autres pourront paraître plus discutables.

Dans cet article, nous faisons l'hypothèse que l'ensemble des membres d'un SIEF détient l'intégralité des données disponibles pour démontrer l'innocuité de la substance considérée. Par conséquent, nous ne discutons pas directement du partage du coût de production des éventuelles données supplémentaires que le SIEF devrait produire[6]. Ce faisant, les règles de partage des coûts que nous proposons sont des mécanismes de compensation dans le sens suivant. Le coût de production d'une donnée a été financé avant

[6]Les axiomes d'additivité et de monotonie par rapport aux données possédées présentés dans cette section discutent partiellement de ce type de problème.

la constitution du SIEF par un ou plusieurs de ses membres. L'échange de la donnée avec d'autres membres du SIEF peut donc donner lieu à une compensation des membres non propriétaires de la donnée envers les membres propriétaires. L'idée sous-jacente à ce principe de compensation est que le flux monétaire des compensations doit s'annuler : les sommes versées par les membres du SIEF désirant accéder à des données sont redistribuées en intégralité aux membres possédant les données. Dans la section 4, l'*axiome de compensation* mettra en lumière cette propriété.

Une deuxième propriété émerge de la description du principe de compensation. La négociation des membres du SIEF pour le partage du coût de production des données peut se dérouler de plusieurs manières différentes. D'un coté, la discussion peut porter en une seule fois sur l'ensemble des données à la disposition du SIEF. D'un autre coté, l'ensemble des données sur lequel porte la négociation peut tout aussi bien être fragmenté en plusieurs groupes. Dans ce deuxième cas, les discussions se succèdent : une négociation concernant un groupe de données a lieu dès que le SIEF a trouvé un consensus à propos du groupe de données précédent. Au final, le traitement d'un membre du SIEF est obtenu en totalisant les résultats de ces différentes négociations. Il semble de bon sens que la somme des compensations effectuées lorsque le partage se fait de manière fragmentée soit égale à la compensation effectuée lorsque le partage se fait en une seule fois. Dans la section 4, nous appellerons cette propriété l'*axiome d'additivité*.

Une autre propriété qui nous semble indiscutable est que deux membres d'un SIEF possédant les mêmes données soient traités de la même manière à l'issue de la négociation. Dans la section 4, l'*axiome de symétrie* imposera un traitement identique pour de tels agents. Une variante de cet axiome requiert un traitement identique pour tous les membres du SIEF lorsque chacun d'entre eux possède le même ensemble de données. Cet autre axiome sera appelé l'*axiome de symétrie faible* dans la section 4.

Une substance pourra faire l'objet de nouvelles études si son innocuité est remise en cause. La controverse sur la toxicité du désherbant Roundup est un exemple marquant de ce phénomène (voir Nelson et Bullock [2003]). Dans un tel cas, les membres d'un SIEF seront sans doute appelés à renégocier le coût du partage des (nouvelles) données par rapport à une situation initiale. On pourrait imaginer que les membres du SIEF fournissant de nouvelles données ne devraient pas être moins bien traités à l'issue de la renégociation qu'avec le mécanisme de compensation de la situation initiale. C'est l'idée de l'*axiome de monotonie par rapport aux données possédées* que nous définirons dans la section 4.

Certaines données peuvent être partagées par plusieurs membres d'un même SIEF. Au cours de la négociation, il est envisageable que les multiples propriétaires d'une donnée soient en concurrence afin de recevoir une compensation pour cette donnée. Par un simple raisonnement économique, on comprend que si un propriétaire exige une compensation positive en échange de l'accès à sa donnée, rien n'empêche un autre propriétaire de cette donnée de proposer un tarif légèrement inférieur. Ce faisant il récupérerait les compensations de tous les membres du SIEF désirant accéder à la donnée concernée. Les firmes d'un oligopole de Bertrand rencontrent les mêmes incitations au moment de fixer les prix de leurs produits. En réitérant ce raisonnement, on conclut que la compensation requise pour accéder à une donnée possédée par plusieurs membres d'un SIEF devrait être nulle. Dans la section 4, nous utiliserons l'*axiome d'insensibilité aux données non exclusives* pour prendre en compte cette idée.

Comme nous l'avons déjà évoqué, la présence de données à échanger par l'intermédiaire du SIEF signifie que des investissements ont déjà été réalisés afin de produire ces données. Il nous semble important que le mécanisme de compensation tienne compte de ces investissements. D'une manière générale, il s'agit de refléter l'idée selon laquelle les différences de traitement entre deux membres d'un SIEF à l'issue de la négociation soient expliquées, dans une certaine mesure, par les différences d'investissements réalisés par ces membres antérieurement à la constitution du SIEF. Plusieurs points de vues sont envisageables. Nous en retiendrons trois, reposant sur les deux scénarios suivants.

Selon le premier scénario, l'écart de traitement entre un membre désirant accéder à une donnée et un membre propriétaire de cette donnée doit précisément correspondre au coût de production de la donnée concernée. Autrement dit, la différence de traitement entre les deux agents doit annuler le coût de production supporté par le membre propriétaire de la donnée concernée. Dans la section 4, l'*axiome de différence du coût de possession* mesurera cet écart de traitement. Une variante de cet axiome précisera l'écart de traitement uniquement lorsqu'une donnée est détenue exclusivement par un agent. Dans la section 4, cette variante sera appelée *axiome de différence du coût de possession exclusif*.

Le second scénario traite plus spécifiquement du cas de possession partagée d'une donnée. La concurrence entre les multiples propriétaires d'une donnée sur laquelle repose l'axiome d'insensibilité aux données non exclusives peut ne pas avoir lieu, en particulier si ces propriétaires s'entendent afin de fixer une même compensation en échange de leur donnée. Dans ce cas, il convient de préciser le montant que les agents non propriétaires de la donnée seraient prêts à payer. Il prend en compte le fait que la production d'une telle donnée soit a été réalisée en commun par ses propriétaires, soit aurait dû l'être afin de réduire le coût total de production pour l'ensemble des propriétaires. Ce scénario traduit le fait que les membres non propriétaires de la donnée pourraient légitimement refuser de compenser le coût total des multiples productions d'une même donnée alors que la mise en commun des moyens de production aurait été possible. En adoptant ce point de vue, chaque membre non propriétaire estime que l'investissement moyen d'un producteur de la donnée aurait dû s'élever au coût de production de la donnée divisé par le nombre de ses propriétaires. La différence de traitement entre un membre non propriétaire et un membre propriétaire de cette donnée doit précisément correspondre à cette dernière quantité. Dans la section 4, l'*axiome de différence du coût de possession partagé* reflétera un tel écart de traitement.

Il est intéressant de noter que les conditions imposées par les trois axiomes discutant de la différence de traitement entre deux membres d'un SIEF sont identiques dans le cas particulier où chaque donnée est détenue par un unique propriétaire. Les exemples contenus dans le document officiel de l'ECHA [2007] s'inscrivent dans cette perspective. Nous reviendrons sur cet aspect dans la section 6.

Enfin, nous pourrions également concevoir que les propriétaires d'une donnée trouvent équitable que le montant total investi afin de produire la donnée soit intégralement compensé par l'ensemble des membres non propriétaires en échange de l'accès à cette donnée. En particulier, ce principe ne précisera pas la part de la compensation supportée par chaque membre non propriétaires de la donnée. Cette dernière propriété sera appelée *axiome de compensation pour les données non possédées* dans la section 4.

3 Le formalisme des jeux de partage de données

Avant de pouvoir décrire de manière rigoureuse les propriétés précédentes, nous devons préciser le formalisme avec lequel nous abordons le partage de l'information à l'intérieur d'un SIEF. Soit N un ensemble fini de $n \geq 1$ agent(s) enregistrés dans un SIEF. Un *problème de partage de données* entre les agents de N est décrit, pour chaque $i \in N$, par l'ensemble fini M_i des données détenues par l'agent i. Un agent peut ne posséder aucune donnée. Dans cet article, les labels i et j sont utilisés pour identifier des agents tandis que le label k est utilisé pour identifier une donnée. Chaque sous-ensemble non vide S de N est appelé coalition d'agent(s) et N est appelé la grande coalition. Pour chaque $S \subseteq N$, $S \neq \emptyset$, on note $M_S = \cup_{i \in S} M_i$ l'ensemble de toutes les données détenues par les membres de la coalition S. Pour simplifier les notations, on écrit M au lieu de M_N pour l'ensemble de toutes les données, et M_i au lieu de $M_{\{i\}}$. Pour chaque $i \in N$, E_i désigne le sous-ensemble de M_i exclusif à l'agent i, c'est-à-dire l'ensemble, possiblement vide, des données possédées par i et par aucun autre agent dans $N \backslash \{i\}$. Autrement dit $E_i = M \backslash M_{N \backslash \{i\}}$. L'entier n_k désigne le nombre d'agents dans N possédant la donnée k, c'est-à-dire $n_k = |\{i \in N : k \in M_i\}|$.

Pour chaque donnée $k \in M$, le réel positif $d_k \in \mathbb{R}_+$ mesure le coût de la donnée k. Dehez et Tellone [2008] interprètent la quantité d_k comme le coût nécessaire à la reproduction de la donnée k. Notre formalisme est suffisamment souple pour que le coût d_k puisse refléter davantage que le simple coût de reproduction d'une donnée. En effet, ce coût peut incorporer d'autres caractéristiques, comme le temps nécessaire à la reproduction ou le fait que la reproduction de la donnée n'est plus possible. A titre d'exemple, si une expérimentation a été réalisée par le passé sur une espèce animale désormais classée en voie de de disparition, alors le coût de la donnée doit intégrer dans une certaine mesure l'impossibilité de répéter l'expérimentation. Le coût de reproduction peut aussi correspondre au coût de production de la donnée, en particulier si on suppose que la reproduction d'une donnée nécessite un investissement identique à celui de sa production originelle.

Un *jeu coopératif à utilité transférable de partage de données* (simplement jeu par la suite) associé à M est une paire (N, C^M) où N est un ensemble d'agents et où la fonction $C^M : 2^N \longrightarrow \mathbb{R}_+$ associe à chaque coalition $S \subseteq N$ le coût $C^M(S) \in \mathbb{R}_+$ de reproduction des données de M non possédées par les membres de S. Formellement, pour chaque $S \subseteq N$, $S \neq \emptyset$, on a

$$C^M(S) = \sum_{k \in M \backslash M_S} d_k$$

et, par convention, $C^M(\emptyset) = 0$. Pour chaque paire de coalitions S et T telles que $S \subseteq T$, on a $C^M(S) \geq C^M(T)$, c'est-à-dire que la fonction C^M est monotone décroissante. Nous désignons par $\mathcal{G}(N)$ l'ensemble de tous les jeux avec l'ensemble d'agents N. Ainsi, nous pouvons adopter la notation $C^M \in \mathcal{G}(N)$ au lieu de $(N, C^M) \in \mathcal{G}(N)$ sans risque de confusion.

Pour chaque jeu $C^M \in \mathcal{G}(N)$ et chaque donnée $k \in M$, construisons le *jeu élémentaire* $C_k^M \in \mathcal{G}(N)$ associé à la donnée k tel que, pour chaque $S \subseteq N$,

$$C_k^M(S) = \begin{cases} d_k & \text{si } k \notin M_S, \\ 0 & \text{si } k \in M_S. \end{cases}$$

Remarquons que pour chaque $S \subseteq N$, on a

$$C^M(S) = \sum_{k \in M} C_k^M(S),$$

ce qui signifie que chaque jeu se décompose comme une somme de jeux élémentaires. Enfin, un *mécanisme de compensation* est une application $f : \mathcal{G}(N) \longrightarrow \mathbb{R}^n$ qui associe à chaque jeu $C^M \in \mathcal{G}(N)$ un vecteur de compensations $f(C^M) \in \mathbb{R}^n$. Si $f_i(C^M) > 0$, l'agent $i \in N$ paie une compensation, alors que si $f_i(C^M) < 0$, il reçoit une compensation.

4 Axiomes

Considérons un mécanisme de compensation f défini pour la classe $\mathcal{G}(N)$. Dans cette section, nous définissons sous la forme d'axiomes pour f les propriétés envisagées dans la section 2. Nous ne reviendrons pas dans le détail sur l'interprétation des axiomes fournie dans la section 2.

Compensation. Pour chaque $C^M \in \mathcal{G}(N)$, on a $\sum_{i \in N} f_i(C^M) = 0$.

L'axiome de compensation pour la classe $\mathcal{G}(N)$ est une réécriture de l'axiome standard d'efficience pour la classe des jeux coopératifs à utilité transférable dans le cas particulier où la valeur de la grande coalition est nulle.

A partir de deux jeux $C^M \in \mathcal{G}(N)$ et $C^W \in \mathcal{G}(N)$ tels que $M \cap W = \emptyset$, construisons le jeu $C^{M \cup W} \in \mathcal{G}(N)$ tel que pour chaque $S \subseteq N$, on a $C^{M \cup W}(S) = C^M(S) + C^W(S)$.

Additivité. Pour chaque jeu $C^M \in \mathcal{G}(N)$ et chaque jeu $C^W \in \mathcal{G}(N)$, on a $f(C^{M \cup W}) = f(C^M) + f(C^W)$.

Deux agents i et j sont *symétriques* dans un jeu C^M si $M_i = M_j$.

Symétrie. Pour chaque jeu $C^M \in \mathcal{G}(N)$ et deux agents i et j symétriques dans C^M, on a $f_i(C^M) = f_j(C^M)$.

Un jeu $C^M \in \mathcal{G}(N)$ est un *jeu symétrique* si pour chaque $i, j \in N$, les agents i et j sont symétriques.

Symétrie faible. Pour chaque jeu symétrique $C^M \in \mathcal{G}(N)$, il existe $b \in \mathbb{R}$ tel que, pour chaque $i \in N$, on a $f_i(C^M) = b$.

Un jeu $C^M \in \mathcal{G}(N)$ est une *extension* d'un jeu $C^W \in \mathcal{G}(N)$ si, pour chaque $i \in N$, on a $M_i \supseteq W_i$.

Monotonie par rapport aux données possédées. Pour chaque jeu $C^M \in \mathcal{G}(N)$, chaque jeu $C^W \in \mathcal{G}(N)$ dont C^M est une extension et chaque $i \in N$, on a $f_i(C^M) \leq f_i(C^W)$.

Deux jeux $C^M \in \mathcal{G}(N)$ et $C^W \in \mathcal{G}(N)$ sont *exclusivement équivalents* si les ensembles des données exclusives de chaque agent sont identiques dans les deux jeux.

Insensibilité aux données non exclusives. Pour deux jeux exclusivement équivalents $C^M \in \mathcal{G}(N)$ et $C^W \in \mathcal{G}(N)$, on a $f(C^M) = f(C^W)$.

Un jeu $C^M \in \mathcal{G}(N)$ est *totalement discriminant* si $M \neq \emptyset$ et si pour chaque $i \in N$, $M_i \in \{\emptyset, M\}$. Dans un jeu totalement discriminant, les agents sont divisés en deux catégories : les propriétaires et les non propriétaires, sachant que chaque propriétaire possède l'ensemble de toutes les données. Un jeu C^M est *exclusif* s'il est totalement discriminant et s'il existe un unique agent $i \in N$ tel que $M_i = M$. Dans un jeu totalement discriminant, un agent possède soit la totalité des données soit aucune donnée. Dans un jeu exclusif, un unique propriétaire détient exclusivement toutes les données.

Les trois axiomes suivants concernent les jeux totalement discriminants. Ils traitent de l'écart de compensation entre les agents non propriétaires et les agents propriétaires. Cet écart est formulé en fonction de la richesse initiale dont disposent les agents propriétaires, mesurée par le coût total des données possédées par chacun de ces agents. Puisque ces données ont été produites antérieurement à la constitution de la grande coalition, les agents non propriétaires peuvent envisager la compensation du coût de ces données de deux manière différentes :

(a) chaque propriétaire (ou l'unique propriétaire si $n_k = 1$ pour chaque $k \in M$) de l'ensemble M des données a lui-même produit chacune des données. Dans ce premier cas, les agents non propriétaires estiment que le coût que chaque propriétaire à dû supporter avant la constitution de la grande coalition est donné par $\sum_{k \in M} d_k$,

(b) avant la constitution de la grande coalition, les $n_k \geq 2$ propriétaires de chacune des données de M se sont regroupés afin de partager les coûts de la production de l'ensemble des données M ou auraient dû le faire afin de minimiser les dépenses pour produire ces données. Dans ce deuxième cas, les agents non propriétaires estiment qu'un coût moyen de $\sum_{k \in M} d_k/n_k$ par propriétaire aurait été suffisant afin de produire les données de M.

Les trois axiomes que nous proposons indiquent que la différence de traitement entre un agent non propriétaire et un agent propriétaire doit refléter précisément les coûts estimés dans les approches *(a)* et *(b)*. Selon l'approche retenue, on obtient deux axiomes intitulés axiome de la différence du coût de possession et axiome de la différence du coût de possession partagé respectivement. Le troisième axiome est une version affaiblie du premier puisque l'écart de traitement est requis uniquement pour les jeux exclusifs.

Différence du coût de possession. Pour chaque jeu totalement discriminant $C^M \in \mathcal{G}(N)$, chaque $i \in N$ tel que $M_i = \emptyset$ et chaque $j \in N$ tel que $M_j = M$, on a $f_i(C^M) - f_j(C^M) = \sum_{k \in M} d_k$.

Différence du coût de possession partagé. Pour chaque jeu totalement discriminant $C^M \in \mathcal{G}(N)$, chaque $i \in N$ tel que $M_i = \emptyset$ et chaque $j \in N$ tel que $M_j = M$, on a $f_i(C^M) - f_j(C^M) = \sum_{k \in M} d_k/n_k$.

Différence du coût de possession exclusif. Pour chaque jeu exclusif $C^M \in \mathcal{G}(N)$, chaque $i \in N$ tel que $M_i = \emptyset$ et chaque $j \in N$ tel que $M_j = M$, on a $f_i(C^M) - f_j(C^M) = \sum_{k \in M} d_k$.

Enfin, l'axiome de compensation pour les données non possédées traduit le fait que le total des compensations versées par les agents non propriétaires doit coïncider exactement avec le coût des données possédées par les propriétaires. Ceci signifie que pour accéder aux données de M, le groupe des agents non propriétaire doit payer exactement la somme qui aurait été nécessaire pour reproduire les données de M.

Compensation pour les données non possédées. Pour chaque jeu totalement discriminant $C^M \in \mathcal{G}(N)$, on a $\sum_{i\in N:M_i=\emptyset} f_i(C^M) = \sum_{k\in M} d_k$.

5 Présentation de cinq mécanismes de compensation

Dans cet article, nous considérons cinq mécanismes de compensation. Le premier est appelé partage égalitaire dans la littérature sur les jeux coopératifs à utilité transférable, et consiste à répartir en n parts égales le coût associé à la grande coalition. Formellement, le *partage égalitaire* est le mécanisme de compensation PE pour la classe $\mathcal{G}(N)$ qui associe à chaque $C^M \in \mathcal{G}(N)$ et chaque $i \in N$ la compensation

$$\mathrm{PE}_i(C^M) = \frac{1}{n} C^M(N).$$

Un deuxième mécanisme de compensation proche du partage égalitaire est le partage égalitaire du surplus, qui consiste pour chaque agent à payer dans un premier temps le coût de faire cavalier seul $C^M(\{i\})$, puis à redistribuer en parts égales le total déjà payé de la sorte par les n agents. Formellement, le *partage égalitaire du surplus* est le mécanisme de compensation PES pour la classe $\mathcal{G}(N)$ qui associe à chaque $C^M \in \mathcal{G}(N)$ et chaque $i \in N$ la compensation

$$\mathrm{PES}_i(C^M) = C^M(\{i\}) + \frac{1}{n}\Big(C^M(N) - \sum_{j\in N} C^M(\{j\})\Big). \tag{1}$$

Le principe du troisième mécanisme de compensation est similaire à celui du partage égalitaire du surplus. Ce troisième mécanisme de compensation, appelé partage égalitaire des coûts marginaux, est défini par une variante de l'expression (1) lorsqu'on substitue à $C^M(\{i\})$ le coût marginal de i à la grande coalition donné par $m_i(C^M) = C^M(N) - C^M(N\backslash\{i\})$. Le *partage égalitaire des coûts marginaux* est le mécanisme de compensation PECM pour la classe $\mathcal{G}(N)$ qui associe à chaque $C^M \in \mathcal{G}(N)$ et chaque $i \in N$ la compensation

$$\mathrm{PECM}_i(C^M) = m_i(C^M) + \frac{1}{n}\Big(C^M(N) - \sum_{j\in N} m_j(C^M)\Big). \tag{2}$$

Le quatrième mécanisme de compensation que nous étudions est la valeur de Shapley introduite par Shapley [1953] pour les jeux coopératifs à utilité transférable. Pour chaque coalition d'agents S qui contient l'agent $i \in N$, la quantité $C^M(S) - C^M(S\backslash\{i\})$ mesure le coût marginal de l'agent i à la coalition S, c'est-à-dire le coût supplémentaire des données que la coalition $S\backslash\{i\}$ doit reproduire si l'agent i la quitte en emportant les données qu'il possède. La valeur de Shapley attribue à chaque agent la moyenne de leur coût marginal aux coalitions du jeu. Formellement, la *valeur de Shapley* est le mécanisme de compensation Sh pour la classe $\mathcal{G}(N)$ qui associe à chaque $C^M \in \mathcal{G}(N)$ et chaque $i \in N$ la compensation

$$\mathrm{Sh}_i(C^M) = \sum_{S \subseteq N : i \in S} \frac{(n-s)!(s-1)!}{n!} \big(C^M(S) - C^M(S \backslash \{i\})\big) \tag{3}$$

où s est le nombre des membres d'une coalition S.

Les solutions proposées par la théorie des jeux coopératifs sont utilisées avec succès par les praticiens. Young [1994] propose une revue de littérature de ces applications. A titre d'exemple, la valeur de Shapley et plusieurs règles de répartition égalitaires servent au partage des coûts de maintenance des canaux d'irrigation dans les ranchs du Montana (Aadland et Kolpin [1998]). La plupart des mécanismes présentés jusqu'à présent a déjà été axiomatisée dans le cadre des jeux coopératifs à utilité transférable. Le lecteur est renvoyé à van den Brink [2007] pour des caractérisations du partage égalitaire et du partage égalitaire du surplus, et à Shapley [1953] et Young [1985] pour des caractérisations de la valeur de Shapley.

Enfin, nous introduisons dans cet article un nouveau mécanisme de compensation appelé mécanisme de compensation intégrale. Le principe de ce mécanisme consiste, pour les agents ne possédant pas une donnée, à se regrouper pour partager égalitairement le coût de reproduction de cette donnée. Chacun de ces agents paie une part égale de ce coût de reproduction, la somme totale payée (le coût de la donnée) étant reversée en parts égales à chaque agent détenteur de la donnée. Formellement, le *mécanisme de compensation intégrale* est le mécanisme de compensation CI pour la classe $\mathcal{G}(N)$ qui associe à chaque $C^M \in \mathcal{G}(N)$ et chaque $i \in N$ la compensation

$$\mathrm{CI}_i(C^M) = \sum_{k \notin M_i} \frac{d_k}{n - n_k} - \sum_{k \in M_i} \frac{d_k}{n_k}. \tag{4}$$

6 Caractérisations des cinq mécanismes

Dans cette section, nous utilisons les axiomes présentés dans la section 4 afin de caractériser les cinq mécanismes de compensation introduits dans la section précédente[7].

Proposition 1 *Le partage égalitaire est l'unique mécanisme de compensation pour la classe $\mathcal{G}(N)$ qui satisfait les axiomes de compensation, de symétrie faible et de monotonie par rapport aux données possédées.*

Preuve. Premièrement, il est facile de vérifier que le partage égalitaire PE satisfait les trois axiomes de l'énoncé pour la classe $\mathcal{G}(N)$. Deuxièmement, montrons que ces trois axiomes font émerger un unique mécanisme de compensation. Pour cela, considérons un mécanisme de compensation f pour la classe $\mathcal{G}(N)$ qui satisfait les trois axiomes de l'énoncé. Considérons le jeu C^W associé à l'ensemble de données $W = \emptyset$. Il est clair que pour chaque $i \in N$, on a $W_i = \emptyset$ si bien que le jeu C^W est symétrique. Par l'axiome de symétrie faible, il existe un réel $b \in \mathbb{R}$ tel que $f_i(C^M) = b$ pour chaque $i \in N$. Par l'axiome de compensation, on obtient $b = 0$. Maintenant, considérons n'importe quel

[7]La preuve de l'indépendance logique des axiomes utilisés n'est pas faite, mais est disponible sur demande auprès des auteurs.

ensemble de données non vide M. On a donc $\emptyset = W_i \subseteq M_i$ pour chaque $i \in N$. Ainsi C^M est une extension de C^W. En appliquant l'axiome de monotonie par rapport aux données possédées, nous savons que pour chaque $i \in N$, on a $0 = f_i(C^W) \geq f_i(C^M)$. En combinant ces n inégalités avec l'axiome de compensation, il vient que $f_i(C^M) = 0$ pour chaque $i \in N$, ce qui conclut la preuve. ■

En somme, le partage égalitaire dans un jeu de partage de données peut être vu comme un mécanisme de compensation purement altruiste : chaque agent met gratuitement à disposition des autres membres du SIEF les données qu'il possède.

Dans chacune des trois propositions suivantes, le mécanisme de compensation étudié vérifie les axiomes de compensation, de symétrie faible et d'additivité. Ces trois propositions se distinguent par l'axiome qu'elles utilisent pour mesurer l'écart de traitement entre un agent non propriétaire et un agent propriétaire dans un jeu totalement discriminant. Si l'axiome considéré est celui de la différence du coût de possession, on caractérise le partage égalitaire du surplus. Lorsqu'on utilise l'axiome de la différence du coût de possession exclusif et que l'on ajoute l'axiome d'insensibilité aux données non exclusives, on caractérise le partage égalitaire des coûts marginaux. Enfin, lorsqu'on utilise l'axiome de la différence du coût de possession partagé, on caractérise la valeur de Shapley.

Proposition 2 *Le partage égalitaire du surplus est l'unique mécanisme de compensation pour la classe $\mathcal{G}(N)$ qui satisfait les axiomes de compensation, de symétrie faible d'additivité et de différence du coût de possession.*

Preuve. Premièrement, il est facile de vérifier que le partage égalitaire du surplus PES satisfait les quatre axiomes de l'énoncé pour la classe $\mathcal{G}(N)$. Deuxièmement, considérons $C^M \in \mathcal{G}(N)$ un jeu quelconque et un mécanisme de compensation f pour la classe $\mathcal{G}(N)$ qui satisfait les quatre axiomes. Fixons une donnée $k \in M$ et considérons le jeu élémentaire C_k^M sur la donnée k. Deux cas de figure sont possibles.

Soit $k \in M_i$ pour chaque agent $i \in N$. Dans ce cas, tous les agents possèdent la donnée k, c'est-à-dire que ces agents sont tous symétriques dans C_k^M. En combinant cette observation avec les axiomes de symétrie faible et de compensation, il vient que $f_i(C_k^M) = 0$ pour chaque agent $i \in N$.

Soit il existe un agent $i \in N$ tel que $k \notin M_i$. Dans ce cas, il existe aussi un agent $j \in N \backslash \{i\}$ tel que $k \in M_j$ puisque $k \in M$. Par définition, le jeu C_k^M est totalement discriminant. Nous pouvons donc appliquer l'axiome de différence du coût de possession aux agents i et j. On obtient $f_i(C_k^M) - f_j(C_k^M) = d_k$. De la même manière, pour un autre agent propriétaire r, on obtient $f_i(C_k^M) - f_r(C_k^M) = d_k$. Ces deux inégalités impliquent que $f_j(C_k^M) = f_r(C_k^M)$ pour deux propriétaires j et r de la donnée k. Par le même raisonnement, on montre facilement que le traitement de deux agents ne possédant pas la donnée k est identique. Autrement dit, un mécanisme qui vérifie l'axiome de différence du coût de possession satisfait également l'axiome de symétrie. Notons b la compensation que chaque agent non propriétaire paie ou reçoit dans C_k^M. Grâce à l'axiome de différence du coût de possession, nous savons que le traitement de chaque agent propriétaire dans le jeu C_k^M est de $b - d_k$. Ainsi, avec l'axiome de compensation, on obtient

$$\sum_{i \in N : k \notin M_i} f_i(C_k^M) + \sum_{j \in N : k \notin M_j} f_j(C_k^M) = 0$$

ou, de manière équivalente, $(n - n_k)b + n_k(b - d_k) = 0$, si bien que $b = n_k d_k / n$. Au total, les axiomes de compensation et de différence du coût de possession font émerger un unique mécanisme de compensation dans chaque jeu élémentaire C_k^M tel que $M \neq \emptyset$. Si i ne possède pas la donnée k dans C^M, il paie $f_i(C_k^M) = n_k d_k/n$, alors que si i possède la donnée k dans C^M, il reçoit la compensation $f_i(C_k^M) = b - d_k = n_k d_k/n - d_k$.

Puisque $M = \cup_{k\in M}\{k\}$, l'axiome d'additivité peut être employé afin d'obtenir l'unique mécanisme de compensation

$$f(C^M) = \sum_{k\in M} f(C_k^M)$$

pour le jeu $C^M \in \mathcal{G}(N)$. Comme le jeu C^M avait été choisi arbitrairement dans $\mathcal{G}(N)$, la preuve est complète. ■

Le partage égalitaire du surplus est donc caractérisé par l'axiome de la différence du coût de possession : la compensation de chaque agent prend en compte le coût supporté par chaque agent pour produire des données. Par ailleurs, notons que pour un jeu $C^M \in \mathcal{G}(N)$, le partage égalitaire du surplus d'un agent $i \in N$ peut se réécrire :

$$\mathrm{PES}_i(C^M) = \sum_{k\in M\setminus M_i} d_k - \frac{1}{n}\sum_{k\in M}(n-n_k)d_k.$$

Proposition 3 *Le partage égalitaire des coûts marginaux est l'unique mécanisme de compensation pour la classe $\mathcal{G}(N)$ qui satisfait les axiomes de compensation, de symétrie faible, d'additivité, de différence du coût de possession exclusif et d'insensibilité aux données non exclusives.*

Preuve. Premièrement, il est facile de vérifier que le partage égalitaire des coûts marginaux PECM satisfait les trois premiers axiomes de l'énoncé pour la classe $\mathcal{G}(N)$. Afin de prouver qu'il satisfait également les deux autres axiomes, il suffit de noter que l'équation (2) se réécrit de la manière suivante :

$$\mathrm{PECM}_i(C^M) = -\sum_{k\in E_i} d_k + \frac{1}{n}\sum_{k\in E} d_k$$

où E désigne l'ensemble de toutes les données exclusives dans M. Deuxièmement, considérons $C^M \in \mathcal{G}(N)$ un jeu quelconque et un mécanisme de compensation f pour la classe $\mathcal{G}(N)$ qui satisfait les cinq axiomes. A partir du jeu $C^M \in \mathcal{G}(N)$, construisons le jeu $C^E \in \mathcal{G}(N)$ obtenu à partir des ensembles de données exclusives E_i des agents $i \in N$. Pour le jeu C^E, la preuve d'unicité est similaire à celle de la preuve de la proposition 2. Si $E = \emptyset$, les axiomes de compensation et de symétrie faible impliquent $f_i(C^E) = 0$ pour chaque $i \in N$. Si $E \neq \emptyset$, il suffit de considérer les jeux élémentaires C_k^E tels que k est la donnée exclusive d'un (unique) agent $i \in N$ dans C^M. Le jeu C_k^E ainsi construit est évidemment un jeu exclusif. Comme dans la preuve de la proposition 2, l'axiome de différence du coût de possession exclusif se traduit par un traitement identique pour chacun des $n-1$ agents non propriétaires de la donnée k dans le jeu exclusif C_k^E. Par conséquent, la combinaison des axiomes de compensation et de différence du coût de possession exclusif font émerger l'unique mécanisme de compensation

$$f_i(C_k^E) = \begin{cases} d_k/n & \text{si } k \notin M_i, \\ d_k/n - d_k & \text{si } k \in M_i \end{cases}$$

dans chaque jeu C_k^E tel que $E \neq \emptyset$. Par définition, pour deux agents distincts i et j, on a $E_i \cap E_j = \emptyset$ et $E = \cup_{i\in N} E_i$. Ainsi, l'axiome d'additivité permet d'obtenir l'unique mécanisme de compensation $f_i(C^E) = \sum_{k\in E} f_i(C_k^E)$ pour chaque $i \in N$ dans C^E. Enfin, les jeux C^M et C^E sont exclusivement équivalents. Par conséquent, l'axiome d'insensibilité aux données non exclusives implique l'égalité $f(C^M) = f(C^E)$. Comme désiré, nous avons obtenu un unique vecteur de compensations pour le jeu C^M. ■

Le partage égalitaire des coûts marginaux est caractérisé par l'axiome de différence du coût de possession exclusif auquel on ajoute l'axiome d'insensibilité aux données non exclusives : la compensation d'un agent prend en compte le coût supporté par chaque agent pour produire des données, à condition que ces données soient exclusives.

Proposition 4 *La valeur de Shapley est l'unique mécanisme de compensation pour la classe $\mathcal{G}(N)$ qui satisfait les axiomes de compensation, de symétrie faible, d'additivité et de différence du coût de possession partagé.*

Preuve. Il est connu que la valeur de Shapley satisfait les trois premiers axiomes de l'énoncé. Afin de vérifier qu'elle satisfait l'axiome de différence du coût de possession partagé pour la classe $\mathcal{G}(N)$, le lecteur peut utiliser la réécriture de l'équation (3) de la proposition 3 dans Dehez et Tellone [2008] : pour chaque $C^M \in \mathcal{G}(N)$ et chaque $i \in N$,

$$\mathrm{Sh}_i(C^M) = \frac{1}{n}\sum_{k\in M} d_k - \sum_{k\in M_i} \frac{d_k}{n_k}.$$

Ensuite, considérons $C^M \in \mathcal{G}(N)$ un jeu quelconque et un mécanisme de compensation f pour la classe $\mathcal{G}(N)$ qui satisfait les quatre axiomes. La preuve est similaire à celle de la proposition 2. Si $M = \emptyset$, on utilise une nouvelle fois les axiomes de compensation et de symétrie faible afin d'obtenir l'égalité $f_i(C^M) = 0$ pour chaque $i \in N$. Si $M \neq \emptyset$, on fixe une donnée $k \in M$ et on considère le jeu élémentaire C_k^M associé à k. Comme dans la preuve de la proposition 2, dans le jeu élémentaire C_k^M, l'axiome de différence du coût de possession exclusif entraîne un traitement identique pour chaque agent non propriétaire de la donnée k, et un traitement identique pour chaque agent propriétaire de cette donnée. Le seul changement par rapport au mécanisme caractérisé dans la proposition 2 est l'ampleur de l'écart de traitement entre un agent non propriétaire et un agent propriétaire. La combinaison des axiomes de compensation et de différence du coût de possession partagé conduit donc à l'unique mécanisme de compensation

$$f_i(C_k^M) = \begin{cases} d_k/n & \text{si } k \notin M_i, \\ d_k/n - d_k/n_k & \text{si } k \in M_i, \end{cases}$$

pour chaque jeu élémentaire C_k^M tel que $M \neq \emptyset$. En utilisant l'axiome d'additivité, on obtient un unique mécanisme de compensation pour C^M. ■

La valeur de Shapley est caractérisée par l'axiome de différence du coût de possession partagé : la compensation d'un agent prend en compte l'apport des données de l'agent dans toutes les combinaisons de coopération possibles au sein de la grande coalition.

Dehez et Tellone [2008] considèrent les jeux de partition. Un jeu C^M est un *jeu de partition* si pour chaque $i, j \in N$, $i \neq j$, on a $M_i \cap M_j = \emptyset$. Autrement dit, chaque agent détient l'exclusivité sur chacune des données qu'il possède. Il est intéressant d'observer que pour cette catégorie de jeux, les axiomes de la différence du coût de possession, de différence du coût de possession exclusif et de différence du coût de possession partagé sont équivalents. De plus, la condition de l'axiome d'insensibilité aux données non exclusives est automatiquement satisfaite pour les jeux de partition. Par conséquent, pour la classe des jeux de partition, le partage égalitaire du surplus, le partage égalitaire des coûts marginaux et la valeur de Shapley coïncident[8].

Corollaire 1 *Soit $C^M \in \mathcal{G}(N)$ un jeu de partition, alors $PES(C^M) = PECM(C^M) = Sh(C^M)$.*

Le dernier résultat de cette section concerne le mécanisme de compensation intégrale.

Proposition 5 *Le mécanisme de compensation intégrale est l'unique mécanisme de compensation pour la classe $\mathcal{G}(N)$ qui satisfait les axiomes de compensation, de symétrie, d'additivité et de compensation pour les données non possédées.*

Preuve. Premièrement, il est facile de vérifier que le mécanisme de compensation intégrale CI satisfait les quatre axiomes de l'énoncé pour la classe $\mathcal{G}(N)$. Deuxièmement, considérons $C^M \in \mathcal{G}(N)$ un jeu quelconque et un mécanisme de compensation f pour la classe $\mathcal{G}(N)$ qui satisfait les quatre axiomes. Fixons une donnée $k \in M$ et considérons le jeu élémentaire C_k^M sur la donnée k. Deux cas de figures sont possibles

Soit $k \in M_i$ pour chaque agent $i \in N$, et dans ce cas tous les agents sont symétriques dans C_k^M. Avec les axiomes de compensation et de symétrie, il vient que $f_i(C_k^M) = 0$ pour chaque agent $i \in N$.

Soit il existe un agent $i \in N$ tel que $k \notin M_i$. Dans ce cas, il existe aussi un agent $j \in N \backslash \{i\}$ tel que $k \in M_j$ puisque $k \in M$. Le jeu C_k^M étant totalement discriminant, nous pouvons appliquer l'axiome de compensation pour les données non possédées. Il vient que

$$\sum_{i \in N : k \notin M_i} f_i(C_k^M) = d_k.$$

Par ailleurs, les agents non propriétaires de la donnée k sont symétriques dans ce jeu élémentaire. L'axiome de symétrie implique donc que $f_i(C_k^M) = d_k/(n - n_k)$ pour chacun de ces agents. Par l'axiome de compensation, nous savons que pour les agents propriétaires de la donnée k, on a

$$\sum_{i \in N : k \in M_i} f_i(C_k^M) = -d_k.$$

Puisque ces agents sont aussi symétriques dans C_k^M, l'axiome de symétrie permet de conclure que $f_i(C_k^M) = -d_k/n_k$ pour chacun de ces agents. Les axiomes compensation, de symétrie et de compensation pour les données non possédées nous ont donc permis d'obtenir un unique mécanisme de compensation dans le jeu élémentaire C_k^M. Par l'axiome d'additivité, on obtient un unique mécanisme de compensation dans le jeu C^M. ■

7 Un exemple numérique

Considérons un SIEF constitué de l'ensemble d'agents $N = \{1, 2, 3\}$. Posons $M_1 = \{w, x, y\}$, $M_2 = \{w, x\}$ et $M_3 = \{y, z\}$ de sorte que $M = \{w, x, y, z\}$. On suppose les

[8]Pour la classe des jeux de partitions, il existe une équivalence avec deux autres mécanismes que nous ne traitons pas dans cet article. Tout d'abord, notre résultat et le théorème 4 de Ju, Borm et Ruys [2007] impliquent que les trois mécanismes concernés coïncident avec la *valeur de consensus*. Ensuite, la proposition 4 de Dehez et Tellone [2008] montre que la valeur de Shapley d'un jeu de partition coïncide avec le *nucléole* (voir Schmeidler [1969] de ce jeu.

coûts suivants : $d_w = 1$, $d_x = 3$, $d_y = 6$ et $d_z = 2$. On obtient la fonction C^M décrite par le tableau 1.

Tableau 1 : Fonction de coût.

S	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
$C^M(S)$	2	8	4	2	0	0	0

Les traitements prescrits par les cinq mécanismes de compensation considérés dans cet article sont répertoriés dans le tableau 2.

Tableau 2 : Compensations des cinq mécanismes.

	Agent 1	Agent 2	Agent 3
Partage égalitaire	0	0	0
Partage égalitaire du surplus	$-8/3$	$10/3$	$-2/3$
Partage égalitaire des coûts marginaux	$2/3$	$2/3$	$-4/3$
Valeur de Shapley	-1	2	-1
Mécanisme de compensation intégrale	-4	5	-1

Les résultats obtenus sont sensiblement différents selon le mécanisme de compensation considéré. Plusieurs commentaires sont possibles. Tout d'abord, le traitement de l'agent 2 est positif dans les cinq mécanismes. Cet agent doit donc payer pour accéder aux données des autres agents, ce qui traduit bien le fait que l'ensemble des données détenu par l'agent 2 a moins de valeur que ceux des autres agents. Ensuite, l'agent 1 perçoit la plus grande compensation sur quatre des cinq mécanismes, ce qui reflète sa possession d'un ensemble de données à la fois volumineux et coûteux. En revanche, le traitement de cet agent est positif sur le partage égalitaire des coûts marginaux car cet agent ne possède aucune donnée exclusive. Sur ce mécanisme, les agents 1 et 2 offrent une compensation à l'agent 3, le seul propriétaire d'une donnée exclusive.

Les résultats de cet exemple numérique illustrent plusieurs propriétés générales des mécanismes et de leur positionnement relatif. Comme nous l'avons déjà évoqué, le partage égalitaire doit être vu comme une procédure altruiste de partage de l'information dans la mesure où chaque agent cède l'accès à ses données sans recevoir aucune compensation. En ce sens, il diffère radicalement des mécanismes du partage égalitaire du surplus, de la valeur de Shapley et du mécanisme de compensation intégrale, qui proposent des compensations similaires mais qui varient dans leur ampleur.

La différence d'ampleur entre le partage égalitaire du surplus et la valeur de Shapley s'explique par les interprétations des investissements réalisés avant la constitution de la grande coalition, selon que l'on retienne intégralement les coûts supportés pour produire les données (partage égalitaire du surplus) ou au contraire que l'on considère les coopérations potentielles ou effectives qui auraient pu réduire le coût de ces investissements (valeur de Shapley). Plus précisément, la compensation reçue par un agent pour chaque donnée k qu'il possède est n_k fois plus importante sur le partage égalitaire du surplus que sur la valeur de Shapley. Sur ce dernier mécanisme, les agents non propriétaires d'une donnée ont conscience que, plus le nombre des propriétaires de cette donnée est élevé, plus leur coopération aurait permis de réduire leur investissement moyen, et donc moins la compensation consentie pour l'accès à cette donnée est élevée.

Dans cet exemple, les compensations perçues par les agents 1 et 3 sont plus importantes sur le mécanisme de compensation intégrale que sur le partage égalitaire du surplus. Ce n'est pas systématiquement le cas, en particulier si le nombre des membres du SIEF est plus élevé. En effet, pour chaque donnée k, la compensation (payée ou reçue) d'un agent sur le mécanisme de compensation intégrale est multipliée par $n/(n_k(n-n_k))$ sur le partage égalitaire du surplus. Prenons l'exemple d'un SIEF contenant $n = 6$ membres. Si la donnée k est possédée par un unique agent (ou par 5 des 6 membres du SIEF), alors la compensation perçue par un propriétaire de k sur le mécanisme de compensation intégrale représente 6/5 de la compensation qu'il reçoit sur le partage égalitaire du surplus. Si la donnée k est possédée par $n_k = 3$ agents, alors la compensation perçue par un propriétaire de k sur le mécanisme de compensation intégrale ne représente plus que 2/3 de la compensation qu'il reçoit sur le partage égalitaire du surplus.

8 Conclusion

Il est utile de préciser que les cinq mécanismes de compensations proposés dans cet article satisfont les axiomes basiques de compensation, de symétrie, de symétrie faible et d'additivité, même si ces axiomes ne sont pas nécessairement utilisés pour leurs caractérisations dans la section 6. Afin d'être complet, le tableau ci-dessous indique les axiomes vérifiés par chacun des cinq mécanismes de compensation.

TABLEAU 3 : Les axiomes et les mécanismes.

	PE	PES	PECM	Sh	CI
Compensation	•	•	•	•	•
Symétrie	•	•	•	•	•
Symétrie faible	•	•	•	•	•
Additivité	•	•	•	•	•
Monotonie par rapport aux données possédées	•				
Différence du coût de possession		•			
Différence du coût de possession partagé				•	
Différence du coût de possession exclusif		•	•	•	
Insensibilité aux données non exclusives	•		•		
Compensation pour les données non possédées					•

Dans un document officiel de l'ECHA [2007], certains mécanismes de compensation sont proposés à titre indicatif. L'un d'eux peut être décrit de la manière suivante. Pour chaque $i \in N$, notons $d(i)$ le coût de la donnée la plus chère possédée par l'agent $i \in N$. Dans l'exemple issu du document de l'ECHA, le coût de (re)production d'une donnée est toujours proportionnelle à sa qualité. Notons également $\bar{d} = \max_{i\in N} d(i)$ et $d = \sum_{i\in N} d(i)$. Sur le mécanisme de compensation proposé dans le document de l'ECHA, chaque agent paie dans un premier temps une fraction $1/n$ de la différence entre le coût de la donnée la plus chère du SIEF et le coût de la donnée la plus chère qu'il possède. Ensuite, il reçoit une part, proportionnelle au coût de la donnée la plus chère qu'il possède, de la somme totale payée par l'ensemble des agents. Formellement, le *mécanisme de compensation Klimisch* associe, à chaque jeu $C^M \in \mathcal{G}(N)$ et à chaque $i \in N$, la compensation

$$\mathrm{CK}_i(C^M) = \frac{\bar{d} - d(i)}{n} - \frac{d(i)}{d} \sum_{j \in N} \frac{\bar{d} - d(j)}{n}.$$

Dans l'exemple numérique de la section 7, on obtient $\mathrm{CK}_1(C^M) = -2/5$, $\mathrm{CK}_2(C^M) = 4/5$ et $\mathrm{CK}_3(C^M) = -2/5$. Un des avantages de ce mécanisme est qu'il incite les agents à produire des données de grande qualité. Par ailleurs, ce mécanisme indique qu'un agent ne possédant aucune donnée paie seulement une fraction $1/n$ de la donnée la plus chère disponible dans le SIEF. Parmi les quatre axiomes vérifiés par les cinq mécanismes de compensation étudiés dans cet article, le mécanisme de compensation Klimisch satisfait les axiomes de compensation, de symétrie, de symétrie faible.

En revanche, ce mécanisme présente de notre point de vue au moins trois lacunes. Tout d'abord, il ne satisfait pas l'axiome d'additivité, ce qui peut s'avérer gênant si de nouvelles données arrivent à disposition du SIEF après un premier partage de données. Avec le mécanisme de compensation Klimisch, il n'est pas garanti que la somme des compensations effectuées lors des deux partages coïncide avec la compensation qui aurait été prescrite si toutes les données avaient été disponibles dès le départ. Ensuite, la compensation payée ou reçue par un agent dépend uniquement, parmi les données qu'il possède, de celle ayant le coût de reproduction (ou la qualité) le plus élevé. Autrement dit, un agent possédant une unique donnée sera traité de la même manière qu'un agent possédant plusieurs données mais dont la plus chère a le même coût de reproduction que l'unique donnée du premier agent. Enfin, les cinq mécanismes proposés dans cet article satisfont tous la propriété intuitive de *monotonie par rapport au coût d'une donnée*. Cette propriété indique que si le coût d'une donnée est revu à la hausse (le coût de chaque autre donnée restant inchangé), alors les propriétaires de cette donnée ne doivent pas recevoir un plus mauvais traitement. Autrement dit, pour chaque agent, le mécanisme de compensation doit être une fonction non croissante du coût de chaque donnée possédée par cet agent. Le mécanisme de compensation Klimisch viole une telle propriété puisque son expression ne tient pas compte de toutes les données possédées par un agent.

Pour conclure cet article, évoquons quelques caractéristiques de la législation REACH que nous n'avons pas abordées. Premièrement, notre formalisme ne prend pas en compte le fait que le coût de production d'une même donnée a pu être différent pour deux de ses propriétaires. Deuxièmement, les jeux de partage de données n'intègrent pas d'éventuelles gains liés à la possession de certains groupes de données complémentaires. Troisièmement, d'après la législation REACH, un membre d'un SIEF n'a pas l'obligation d'accéder à toutes les données à la disposition du SIEF, en particulier si ce membre n'est pas producteur de la substance. Cet aspect a été laissé de côté dans cet article par soucis de simplicité. Cependant, il est facile de modifier notre modèle pour en tenir compte. Pour cela, il suffit de définir, pour chaque agent $i \in N$, un ensemble X_i des données qu'il souhaite posséder. Cet ensemble contient l'ensemble M_i des données produites par l'agent i et est contenu dans l'ensemble M de toutes les données disponibles dans le SIEF. A l'aide de cette nouvelle collection d'ensembles, nous pouvons définir la fonction $C^{M,X} : 2^N \longrightarrow \mathbb{R}_+$

qui associe à chaque coalition S le coût

$$C^{M,X}(S) = \sum_{k \in X_S \setminus M_S} d_k$$

de reproduction des données désirées par au moins un agent de S mais possédées par aucun agent de S. Quatrièmement, nous n'avons pas discuté de l'impact sur le mécanisme de compensation de l'élargissement d'un SIEF à de nouveaux membres. Cinquièmement, nous avons adopté une approche par la théorie des jeux coopératifs. Le partage de l'information peut également être étudié grâce à la littérature sur le partage des coûts qui s'est développée depuis une vingtaine d'années (voir Moulin et Sprumont [2007]). Dans un travail futur, il serait intéressant de comparer les mécanismes de compensation envisagés dans cet article à certaines méthodes de partages des coûts, en particulier la *répartition séquentielle des coûts* proposée par Moulin et Shenker [1992].

Références bibliographiques

AADLAND D., KOLPIN V. [1998], "Shared irrigation costs: An empirical and axiomatic analysis", Mathematical Social Sciences, 35, p. 203–218.

VAN DEN BRINK R. [2007], "Null or nullifying players: The difference between the Shapley value and equal division solutions", Journal of Economic Theory, 136, p. 767–775.

DEHEZ P., TELLONE D. [2008], "Data games. Sharing public goods with exclusion", CORE discussion paper 2008/10.

ECHA [2007], "Guidance for the implementation of REACH: Guidance on data sharing", document de l'European Chemicals Agency, disponible à l'adresse : `http://guidance.echa.europa.eu/docs/guidance_document/data_sharing_en.pdf`

JU Y., BORM P., RUYS P. [2007], "The consensus value: a new solution concept for cooperative games", Social Choice & Welfare, 28, p. 685–703.

MOULIN H. [1981], Théorie des jeux pour l'économie et la politique, Hermann, Paris.

MOULIN H., SHENKER S. [1992], "Serial cost sharing", Econometrica, 50, p. 1009–1039.

MOULIN, H., SPRUMONT Y. [2007], "Fair allocation of production externalities: recent results", Revue d'Economie Politique, 117, p. 7–36.

NELSON G. C., BULLOCK D. S. [2003], "Simulating a relative environmental effect of glyphosate-resistant soybeans", Ecological Economics, 45, p. 189–202.

SCHMEIDLER D. [1969], "The nucleolus of a characteristic function game", SIAM Journal of Applied Mathematics, 17, p. 1163–1170.

SHAPLEY L. S. [1953], "A value for n-person games", Contributions to the theory of games (Kuhn H.W., A.W. Tucker eds), vol. II, Princeton University Press, 307–317.

THOMSON W. [2001], "On the axiomatic method and its recent applications to game theory and ressource allocation", Social Choice and Welfare, 18, p. 327-386.

VINCENT A. L., GARDERES N. [2008], "REACH, les consortiums et le droit de la concurrence", Bulletin du Droit de l'Environnement Industriel, 14, p. 36–38.

YOUNG H. P. [1985], "Monotonic solutions of cooperative games", International Journal of Game Theory, 14, p. 65–72.

YOUNG H. P. [1994], Equity in theory and practice, Princeton University Press.