

Evaluation multicritère appliquée au bien-être des animaux en ferme ou à l'abattoir : difficultés et solutions du projet Welfare Quality®

I. VEISSIER¹, R. BOTREAU¹, P. PERNY²

¹ INRA, UR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

² Université Pierre et Marie Curie, LIP6, 4 place Jussieu, F-75252 Paris, France

Courriel : isabelle.veissier@clermont.inra.fr

La durabilité des systèmes de production, la qualité environnementale des exploitations, le bien-être des animaux, etc., sont autant de concepts complexes dont l'évaluation doit couvrir de nombreux aspects distincts. Le présent article illustre la thématique de l'évaluation de tels concepts à partir de l'exemple du bien-être animal. Il présente les difficultés rencontrées, à la fois éthiques et techniques, et les solutions trouvées dans le projet européen Welfare Quality® grâce à un large processus de consultation et à l'apport des méthodologies multicritères.

Le bien-être animal demeure une préoccupation sociétale importante pour les citoyens européens (European Commission, 2005, 2007a et b). Bien qu'il ne suffise pas d'évaluer un problème pour le corriger, un dispositif d'évaluation permet de détecter les problèmes de bien-être et de mettre en place des solutions correctrices (Polten 2007). De plus, une évaluation du bien-être des animaux pourrait être assortie d'une information auprès des consommateurs, permettant une sélection des produits respectant le bien-être animal et une plus-value pour les producteurs correspondants (Webster 2009, Napolitano *et al* 2010). Or, les consommateurs européens ne s'estiment pas suffisamment informés sur le niveau de bien-être des animaux, et ne sont pas en mesure de le prendre en compte dans leurs décisions d'achat de produits alimentaires (European Commission 2007a et b). En Europe, plusieurs programmes ont été mis en place afin de garantir aux consommateurs un certain niveau de bien-être des animaux dont ils consomment les produits. Citons, par exemple, le *Freedom Food Scheme* au Royaume-Uni, et l'*IKB* adopté par la filière de la viande aux Pays-Bas. Ces initiatives peuvent varier par les indicateurs utilisés pour contrôler le bien-être des animaux, par les seuils fixés pour définir les niveaux acceptables, ou encore par la manière dont l'information est agrégée pour fournir une appréciation globale. Un outil standardisé d'évaluation semble dès lors nécessaire afin d'assurer la crédibilité des affirmations relatives

à la prise en compte du bien-être. En parallèle, dans son livre blanc au parlement, la Commission Européenne avait lancé l'idée d'un outil standardisé de certification en matière de bien-être animal qui pourrait être utilisé dans le cadre de négociations bilatérales entre pays, d'où la nécessité d'un modèle normatif d'évaluation (Commission Européenne 2002).

Le projet de recherche européen Welfare Quality® (2004-2009), financé par la Commission Européenne au sein du 6^{ème} Programme-Cadre de Recherche et de Développement, visait entre autres à proposer un outil d'évaluation des fermes ou des abattoirs au regard du bien-être des animaux (bovins, porcs ou volailles) qu'ils hébergent, outil pouvant servir de base à la fois au conseil en élevage et à des programmes de certification. Plus de 200 chercheurs ou ingénieurs en sciences biologiques ou sociales, issus d'une quarantaine d'organismes de recherche ou de développement de 13 pays européens participaient à ce projet.

Le bien-être d'un individu correspond à la qualité de vie telle qu'il la perçoit (Bracke *et al* 1999). Il recouvre à la fois la bonne santé, l'absence d'émotions négatives, l'expression des comportements, etc. (Farm Animal Welfare Council 1992, Fraser *et al* 1997, Bracke *et al* 1999). Il s'agit d'un concept multidimensionnel, chaque aspect devant être vérifié indépendamment des autres. L'évaluation globale du

bien-être d'individus nécessite donc des mesures variées pour vérifier chaque aspect et un modèle multicritère permettant d'agréger les informations ainsi générées et de passer de la description de l'état des animaux à une évaluation de la ferme ou de l'abattoir (dénommés ci après *exploitations*). Au sein de Welfare Quality®, un groupe de travail était plus particulièrement chargé de développer un tel modèle. Ce groupe était formé de treize chercheurs biologistes, vétérinaires, éthologistes ou mathématiciens. Les trois auteurs de l'article y tenaient des rôles-clés, P. Perny apportant l'expertise en méthodologie multicritère, R. Botreau concevant le modèle pilote sur vaches laitières et I. Veissier animant le groupe et coordonnant les activités. Les auteurs rapportent ici les difficultés rencontrées par le groupe de travail et les étapes de l'élaboration du modèle d'évaluation. Le présent article résume six articles publiés dans des revues internationales (Botreau *et al* 2007a, b et c, 2008, 2009, Veissier *et al* 2010).

1 / Les difficultés d'une évaluation globale du bien-être des animaux

La première difficulté que nous avons rencontrée tenait dans le fait que le bien-être est un concept multidimensionnel, incluant à la fois la santé physique et mentale, ainsi que divers aspects tels que le confort physique,

l'absence de faim, l'absence de pathologies, la possibilité d'exprimer des comportements normaux, etc. A certains égards, évaluer globalement une exploitation à partir de nombreux indicateurs peut être considéré comme «*additionner des choux et des carottes*», c'est-à-dire des éléments de nature très différente. Aussi une agrégation réalisée de manière non pertinente pourrait-elle avoir pour effet de masquer les problèmes de bien-être.

La deuxième contrainte résidait dans le fait qu'un ensemble d'informations est nécessaire pour caractériser le bien-être des animaux dans un élevage donné (par exemple, dans la ferme A, 2% des vaches sont trop maigres, 5% boitent, 50% des vaches fuient lorsqu'une personne s'approche d'elles à moins d'1 m, ...), et qu'il faut émettre un jugement à partir de ces informations. Or un jugement ne peut être exempt de valeurs. Ce problème devient plus criant lorsqu'un jugement doit être porté non pas à partir d'une seule mesure (par exemple, «*2% de vaches trop maigres*» constitue-t-il un bon ou un mauvais résultat ?), mais à partir d'un groupe de mesures (par exemple, «*2% de vaches trop maigres + 5% de vaches souffrant de boiterie + 15% de vaches craignant l'homme*» peut-il être considéré comme un bon résultat ?). En essayant de construire un modèle d'évaluation globale du bien-être animal, nous avons ainsi été confrontés à des dilemmes éthiques et des difficultés techniques.

1.1 / Dilemmes éthiques

Le bien-être d'un animal est déterminé par les conditions de vie auxquelles il est confronté : conditions de logement, d'alimentation, d'allotement, etc. Lorsqu'on considère un seul aspect du bien-être, le point de vue de l'animal peut être appréhendé, du moins en termes de préférences à court terme (Holm *et al* 2007). Par exemple, il est possible de connaître les préférences alimentaires des animaux. En revanche, il semble impossible de savoir selon quelle hiérarchie un animal classerait plusieurs aspects différents du bien-être. Cela devient irréaliste quand ces aspects interviennent à des échelles de temps différentes, comme le fait de ne pas être dérangé par des événements stressants, et avoir des risques de souffrir d'une maladie chronique. Force est de constater qu'au moins à l'heure actuelle, l'évaluation globale du bien-être d'un animal ne peut, dans une certaine mesure, être réalisée que par l'intermédiaire de l'Homme. Le résultat de cette évaluation pourra donc dépendre de la personne qui la réalise. Le choix des personnes à consulter pour construire un outil d'évaluation du bien-être pose

question. Qui est plus à même de comprendre comment un animal perçoit son environnement ?

Certains arbitrages doivent être effectués afin d'établir la manière dont les différentes dimensions sont agrégées pour construire une estimation globale du bien-être d'un animal. Premièrement, certains d'entre nous peuvent estimer que tels aspects sont plus importants que d'autres, tandis que d'autres les considèrent comme de moindre importance. Par exemple, si certains placent la santé avant le comportement, d'autres privilégient à l'inverse les critères comportementaux (Fraser 1995). Deuxièmement, certains acceptent des compensations entre les différentes mesures du bien-être, alors que d'autres ne les admettent pas. Ainsi, lors des discussions qui ont accompagné l'élaboration des directives européennes sur la protection des veaux, un dilemme est apparu entre autoriser les cases individuelles, qui limitent très fortement le comportement des veaux, ou rendre obligatoire la case collective, qui offre plus de possibilités de comportement mais est accompagnée d'une plus forte mortalité (Broom *et al* 1995). Il a finalement été décidé d'interdire la case individuelle au-delà d'un certain âge car la bonne santé des animaux ne peut justifier la restriction extrême de mouvements imposée par la case individuelle alors que des plans sanitaires peuvent être mis en place pour limiter les maladies.

Un problème similaire, et encore plus aigu, surgit lorsqu'il faut évaluer le bien-être non plus d'un individu, mais d'un groupe d'animaux. Doit-on se concentrer sur les individus dont l'état est le plus dégradé ou prendre en compte un niveau de bien-être «moyen» à l'échelle du groupe, ce qui reviendrait à compenser le mauvais état des uns par le bon état des autres ? A ce stade, il n'est pas réaliste d'attendre d'un groupe d'animaux qu'il fournisse lui-même une appréciation de son bien-être global, par exemple au travers de tests de préférence. Une fois encore, l'évaluation du bien-être général d'un groupe d'animaux est inévitablement fondée sur un jugement humain.

Parallèlement, s'agissant de l'évaluation globale d'une ferme, certains estiment que les critères pris en compte doivent être évalués par rapport aux conditions d'exploitation existantes. Ils comparent les résultats obtenus par une exploitation à ceux des autres exploitations de la même population et considèrent que les meilleures exploitations offrent à leurs animaux un bon niveau de bien-être (Whay *et al* 2003). Pour d'autres, une légère amélioration par

rapport à un niveau actuel est insuffisante pour atteindre un bon niveau de bien-être (Bekoff 2008). Selon cette approche, l'appréciation du bien-être doit être fondée sur ce qu'on considère théoriquement comme excellent, bon, etc.

1.2 / Difficultés techniques

La principale contrainte technique est liée au fait que les informations dont nous disposons pour construire le modèle d'évaluation peuvent être de nature très différentes :

- certaines données sont nominales, par exemple la méthode utilisée pour l'écorchage des jeunes bovins (cautérisation au fer rouge ou application de potasse caustique) ;
- certaines données sont ordinales, c'est-à-dire par catégories graduelles. Par exemple, les porcs à l'engrais peuvent avoir un poids normal, être maigres, ou encore très maigres. Dans ce cas, on ne peut pas déterminer si la différence entre le poids normal et la maigreur est similaire à celle qui sépare la maigreur de la maigreur prononcée ;
- certaines données sont cardinales, par exemple la distance de fuite d'un animal face à une personne qui s'approche de lui, laquelle peut être mesurée en mètres.

Cependant, même dans le cas de données cardinales, les variations dans l'interprétation des résultats ne sont pas forcément proportionnelles à l'écart observé entre les données de base. Ainsi, il est peu probable que l'on porte un jugement différent sur une vache présentant une distance de fuite de 3,5 m par rapport à une autre se laissant approcher jusqu'à 3 m, alors qu'on jugera de manière très distincte une vache ayant une distance de fuite de 0,5 m et une autre se laissant toucher par l'observateur (ce qui correspond à une distance de fuite de 0 m), bien que la différence de distance soit rigoureusement identique (0,5 m).

Toutes les observations doivent par conséquent être transposées sur une même échelle de valeur, reflétant le niveau de bien-être des animaux, par exemple 0 pour une absence totale de bien-être et 100 pour un bien-être maximal.

L'outil d'évaluation devant intégrer un certain nombre de choix éthiques, il s'agissait de concevoir un modèle assez souple pour les incorporer finement. La grande variabilité des problèmes rencontrés dans la pratique a conduit des scientifiques issus de différentes disciplines (sciences de gestion, psychologie, mathématiques, économie, recherche opérationnelle et informatique) à

mettre au point divers modèles et méthodologies formels d'aide à l'évaluation et à la décision (Roy 1985), notamment les méthodes de surclassement fondées sur l'agrégation ordinale (Roy 1991, Perny 1998) et la théorie de l'utilité multi-attribut (MAUT, *Multiattribute Utility Theory* (Bouyssou *et al* 2000)) basée sur l'agrégation cardinale (additive ou non-additive ; par exemple, les intégrales de Choquet (Grabisch et Roubens 2000)). Ces méthodes concernent différents aspects relatifs à l'évaluation multicritère, tels que les problèmes de mesure (comment rendre compte des préférences, des perceptions ou des performances au moyen de données numériques) ou les problèmes d'agrégation (comment obtenir une évaluation globale à partir de points de vue multidimensionnels et éventuellement antagonistes) (Bouyssou 1990, Roy 1993, Bouyssou *et al* 2000). L'ensemble de ce corpus méthodologique est destiné à faciliter l'élaboration scientifique de la décision, et peut être utile à toutes les étapes d'un processus décisionnel. Nous avons exploré ces méthodes afin de produire des outils fiables d'évaluation globale du bien-être animal, et d'aide à la décision dans ce domaine.

2 / Architecture du modèle d'évaluation Welfare Quality®

Nous nous sommes appuyés sur les résultats des *focus groups* de citoyens et des entretiens d'éleveurs conduits par les chercheurs en sciences sociales de Welfare Quality® (Miele *et al* accepté) ainsi que sur les discussions entre les scientifiques de Welfare Quality® pour définir quatre grands principes à respecter afin de garantir un bon niveau de bien-être :

- **Alimentation adaptée** : l'apport en nourriture et en eau des animaux est-il correct ?
- **Logement correct** : les conditions de logement des animaux sont-elles appropriées ?
- **Bonne santé** : l'état sanitaire des animaux est-il satisfaisant ?
- **Comportement approprié** : le comportement des animaux reflète-t-il des états émotionnels positifs ?

Chacun de ces principes se subdivise en 2 à 4 critères devant être pris en compte afin d'obtenir un bon niveau de bien-être (tableau 1).

Pour chaque critère, les exploitations sont évaluées à l'aide d'une ou plusieurs mesures portant sur différents aspects du

Tableau 1. Principes et critères du bien-être animal retenus par le projet Welfare Quality®.

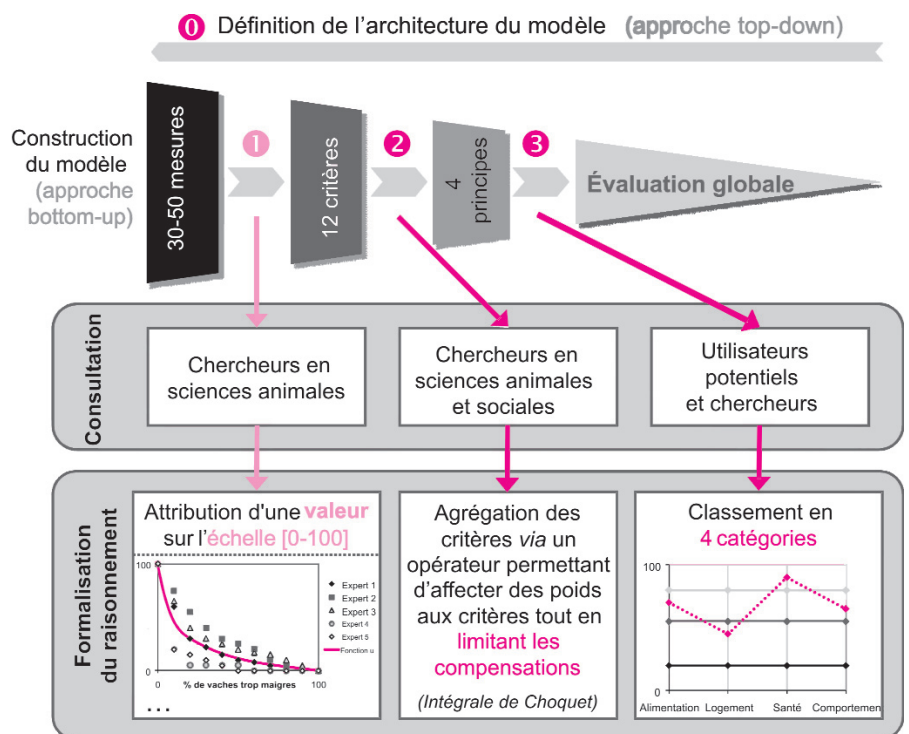
Principes	Critères	
Alimentation adaptée	1	Absence de faim prolongée
	2	Absence de soif prolongée
	3	Confort autour du repos
Logement correct	4	Confort thermique
	5	Facilité de déplacement
	6	Absence de blessures
Bonne santé	7	Absence de maladies
	8	Absence de douleurs causées par les pratiques d'élevage
	9	Expression des comportements sociaux
Comportement approprié	10	Expression des autres comportements
	11	Bonne relation Homme-Animal
	12	Etat émotionnel positif

critère étudié. Ainsi, on contrôle l'absence de maladies en enregistrant la prévalence et la sévérité des signes cliniques les plus courants. De même, on vérifie le confort de la zone de repos des vaches laitières en observant la facilité du coucher et la propreté des animaux. Les mesures retenues sont décrites dans les protocoles finaux de Welfare Quality® (Welfare Quality® 2009a, b et c).

Si les principes et critères retenus pour évaluer les différentes dimensions du bien-être reposent sur une approche

top ⇒ down (définition des principes, puis des critères, et enfin des mesures permettant de vérifier le respect du critère étudié), la construction de l'évaluation globale du bien-être sur une exploitation donnée suit un schéma bottom ⇒ up (figure 1). D'abord, les scores des critères sont calculés à partir des résultats enregistrés pour les différentes mesures sur une exploitation, puis ils sont agrégés afin d'obtenir les scores de principes, ces derniers étant eux-mêmes synthétisés pour produire l'évaluation globale.

Figure 1. Architecture et construction du modèle Welfare Quality® visant à fournir une évaluation globale du bien-être des animaux d'une exploitation.



3 / Les processus de consultation

Au sein de Welfare Quality[®], le groupe de travail était chargé de l'élaboration du modèle d'évaluation du bien-être. Nous avons estimé que cela ne nous autorisait pas à prendre seuls les décisions relatives aux questions éthiques mentionnées précédemment. Nous avons développé un modèle offrant la liberté de prendre, face à une question éthique, telle décision ou telle autre. Nous pouvions ainsi autoriser ou non les compensations entre principes ou critères, donner la priorité aux animaux en moins bon état ou prendre en compte l'état moyen des animaux d'une exploitation. Nous avons ensuite consulté des personnes tierces afin de s'appuyer sur leur avis pour paramétrer le modèle d'évaluation. Le choix du type de personnes consultées a été décidé en concertation avec le comité directeur de Welfare Quality[®] (composé de 6 chercheurs en sciences animales et 2 en sciences sociales) :

– pour la première étape de construction du modèle d'évaluation, c'est-à-dire de la collecte des résultats issus des mesures au calcul des scores pour les différents critères, nous avons consulté les chercheurs en sciences animales qui avaient développé les mesures au sein du projet. Nous avons considéré qu'ils étaient les plus compétents pour appréhender la signification de ces mesures au regard du bien-être des animaux. De plus, ils possédaient une certaine expérience de terrain leur permettant de savoir quels types de résultats pouvaient être obtenus. Quatre à huit chercheurs

ont été consultés selon le critère et l'espèce animale étudiés ;

– pour la deuxième étape, c'est-à-dire le passage des scores de critères aux scores de principes, nous avons consulté de nouveau des chercheurs en sciences animales et également des chercheurs en sciences sociales ayant un rôle déterminant dans Welfare Quality[®] (responsables de groupes de tâches). Les chercheurs en sciences sociales ont fourni le point de vue de la population sur laquelle portaient leurs travaux au sein du projet (citoyens, distributeurs ou producteurs). Les chercheurs en sciences animales ont été considérés comme porteurs du point de vue des animaux. Une quinzaine de personnes ont été consultées ;

– pour l'étape finale, c'est-à-dire le passage des scores au niveau des principes à la production de l'évaluation globale, nous avons consulté les mêmes chercheurs qu'à l'étape précédente ainsi que des porteurs d'enjeux (membres du comité consultatif de Welfare Quality[®] : représentants des organisations d'éleveurs ou sélectionneurs, de distributeurs, de vétérinaires, de la Société Européenne sur l'Éthique en matière d'Agriculture, d'organismes officiels (Office International des Epizooties) et une association de protection des animaux (*Eurogroup for animal*). La contribution de ces porteurs d'enjeux revêtait une importance décisive à ce stade car ils représentaient les utilisateurs potentiels du modèle d'évaluation. Or, le développement du modèle doit être raisonné en fonction du type d'évaluation globale souhaitée, en relation avec la manière dont celle-ci sera utilisée en pratique.

A aucun moment, nous n'avons demandé à ces différents intervenants d'arbitrer des dilemmes éthiques. Nous leur avons simplement soumis des jeux de données, en leur demandant de réagir sur ces informations (exemples présentés dans les tableaux 2 et 3) et d'attribuer des scores sur une échelle de 0 à 100, dans laquelle :

- 0 correspond à la pire situation envisageable au niveau d'une exploitation (c'est-à-dire la situation en deçà de laquelle aucune détérioration supplémentaire du niveau de bien-être n'est possible) ;

- 50 correspond à une situation neutre (le niveau de bien-être n'est ni bon ni mauvais) ;

- 100 correspond à la meilleure situation envisageable (c'est-à-dire la situation au-delà de laquelle aucune amélioration supplémentaire du niveau de bien-être n'est possible).

Par ailleurs, nous avons établi qu'une exploitation présentant un score inférieur à 20 aurait de fortes chances d'être exclue de tout programme de certification visant à garantir le bien-être des animaux. Cette limite de 20 a été validée par le comité directeur de Welfare Quality[®].

Pour l'étape finale de construction du modèle d'évaluation, la consultation s'est déroulée sous la forme de discussions entre scientifiques et porteurs d'enjeux, à l'occasion de réunions et d'échanges de courriers électroniques. Les usages potentiels du système d'évaluation ont été débattus, ainsi que les méthodes d'agrégation des scores obtenus au niveau des quatre grands

Tableau 2. Jeux de données présentés aux chercheurs en sciences animales pour leur permettre d'attribuer un score sur le critère «Absence de blessures», selon la proportion de bovins présentant une blessure légère (par exemple, des pertes de poils) ou une blessure sérieuse (par exemple, une plaie ou un oedème).

Les experts devaient classer les fermes de la pire à la meilleure, puis accorder une note à chacune d'entre elles.

	% de bovins ne présentant pas de blessure	% de bovins présentant une blessure légère	% de bovins présentant une blessure sérieuse	Classement de la ferme	Score final
Ferme 1	100	0	0	?	?
Ferme 2	0	100	0	?	?
Ferme 3	0	0	100	?	?
Ferme 4	50	50	0	?	?
Ferme 5	0	50	50	?	?
Ferme 6	50	0	50	?	?
Ferme 7	50	25	25	?	?
Ferme 8	20	50	30	?	?
Ferme 9	0	25	75	?	?
Ferme 10	75	25	0	?	?
Ferme 11	75	0	25	?	?

Tableau 3. Combinaisons de scores des critères présentées aux chercheurs en sciences animales ou sociales pour leur permettre d'attribuer des scores au niveau des principes.

Score obtenu pour le critère «Absence de faim prolongée»	Score obtenu pour le critère «Absence de soif prolongée»	Score pour le principe «Alimentation adaptée»
25	75	?
40	60	?
50	50	?
60	40	?
75	25	?

principes. Les intervenants ont également pris en compte la répartition des résultats entre les fermes visitées dans le cadre de Welfare Quality®.

4 / Le modèle d'évaluation multicritère

4.1 / Construction des critères à partir des mesures

Les données des mesures se rapportant à un même critère sont interprétées et synthétisées afin de produire un score au niveau du critère, sur l'échelle de 0 à 100 décrite précédemment, indiquant le degré de conformité de l'exploitation à ce critère spécifique.

Le calcul des scores varie en fonction du nombre de mesures, de l'échelle sur laquelle les données sont recueillies et de l'importance relative des mesures intra-critère. Trois principaux modes de calcul sont utilisés :

- Lorsque toutes les mesures servant à examiner un critère sont relevées au niveau de l'exploitation dans son entier et que les résultats sont exprimés en un nombre limité d'alternatives (données nominales), un arbre de décision est construit. Le score d'une alternative correspond au score moyen que les chercheurs interrogés ont attribué à cette alternative. Un exemple est fourni dans l'encadré 1.

- Lorsqu'un critère est évalué à partir d'une seule mesure relevée au niveau

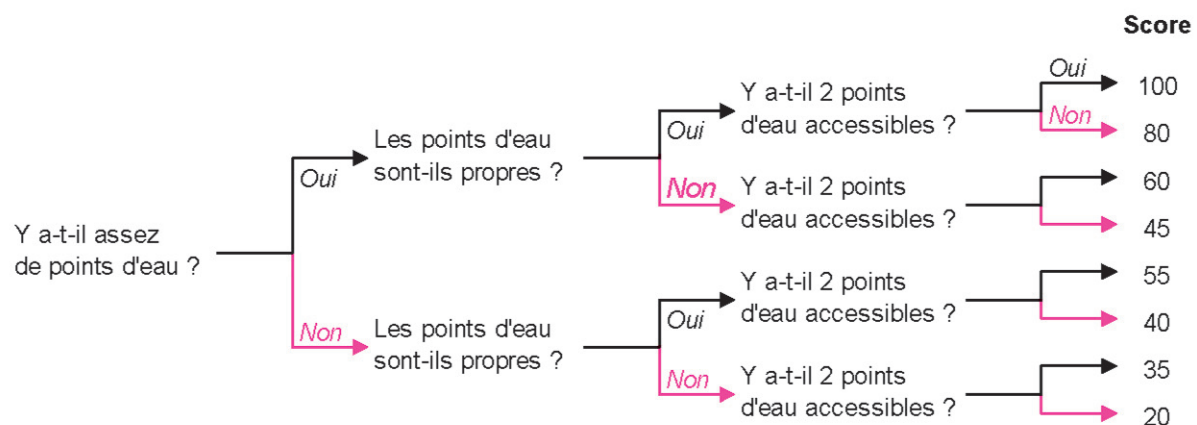
des animaux, mais selon différentes classes (données ordinales ; par exemple, problèmes légers à sérieux), nous calculons la proportion d'animaux observés dans chaque classe (par exemple la proportion d'animaux ne présentant pas de boiterie, ceux présentant une boiterie légère et ceux présentant une boiterie sérieuse). Dans ce cas, une somme pondérée est calculée, avec des poids croissant en fonction de la gravité du problème. Les poids sont choisis de sorte à ce que cette somme reflète le classement des exploitations selon le score moyen attribué par les chercheurs interrogés. Un exemple est fourni dans l'encadré 2.

- Lorsque les mesures servant à évaluer un critère produisent des données exprimées sur différentes échelles (par exemple, pourcentage d'animaux couchés en dehors de la zone prévue à cet effet, temps mis pour se coucher exprimé en secondes), ces données sont comparées à un seuil d'alerte représentant la limite entre situation normale et anormale. La somme des alertes est ensuite calculée. Il est possible de définir une situation intermédiaire (avertissement). Dans ce cas et de façon similaire au paragraphe précédent, c'est une somme pondérée des alertes et avertissements qui est calculée. Un exemple est fourni dans l'encadré 3.

Cet exercice a démontré que les personnes interrogées ne modifient pas

Encadré 1. Arbre de décision construit pour le critère «Absence de soif prolongée» appliqué aux porcs à l'engrais.

La soif n'est pas évaluée directement sur les animaux car les symptômes de déshydratation ne sont détectables que dans les situations extrêmes de privation. Le nombre d'abreuvoirs, leur état de fonctionnement et leur propreté sont observés. Le nombre maximum de porcs recommandé est calculé (10 individus par abreuvoir en état de fonctionnement correct et 5 par abreuvoir de capacité amoindrie). Si le nombre de porcs dans l'enclos est supérieur aux recommandations, le nombre de points d'eau est considéré comme insuffisant. Dans l'arbre de décision ci-après, la propreté des abreuvoirs et l'accès des animaux à au moins deux points d'eau dans l'enclos figurent parmi les sous-critères évalués :



Encadré 2. Somme pondérée et fonctions I-splines appliquées aux vaches laitières présentant des boiteries.

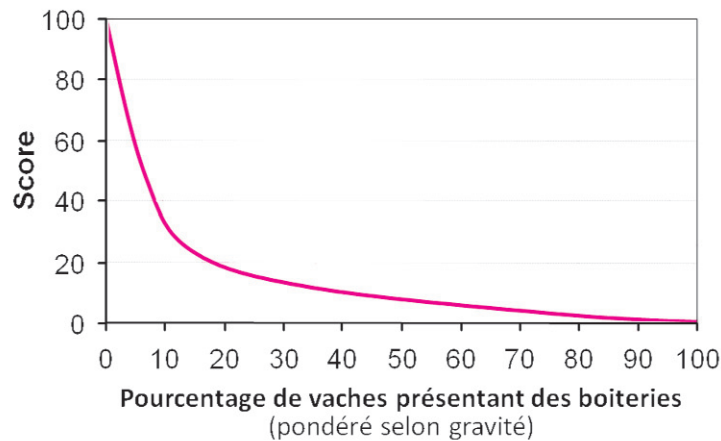
Les pourcentages respectifs d'animaux présentant des boiteries légères et sérieuses sont combinés pour former une somme pondérée, avec un poids de 0,29 pour les premières et de 1 pour les secondes. Cette somme est ensuite convertie en un index variant de 0 à 100 :

$$\text{Index de boiterie } I = 100 - [0,2857 \times (\% \text{ boiteries légères}) + (\% \text{ boiteries sévères})]^*$$

Cet index est transformé en un score au moyen de fonctions I-splines :

$$\text{Lorsque } I \leq 78, \text{ le Score} = (0,09879932469 \times I) - (0,0009549718848 \times I^2) - (5,344391432E^{-05} \times I^3)$$

$$\text{Lorsque } I \geq 78, \text{ le Score} = -2060,598372 + (79,35251951 \times I) - (1,017025859 \times I^2) + (0,004395605013 \times I^3)$$



* L'index de boiterie correspond donc à $100 - \text{le \% de vaches présentant des boiteries (pondéré selon gravité)}$

toujours leur jugement de façon proportionnelle aux variations mesurées sur animaux. Ainsi, le score du critère «absence de blessures» varie peu lorsque la prévalence des boiteries dans un troupeau laitier augmente de 90 à 100% mais il varie beaucoup

lorsque la prévalence augmente de 0 à 10%, alors que dans les deux cas l'augmentation de prévalence est de 10 points (figure de l'encadré 2). Il convient par conséquent de recourir à des fonctions non linéaires pour transformer les données de base en scores

au niveau des critères, en l'occurrence des fonctions I-splines. En bref, les fonctions I-splines permettent de calculer des portions de courbes afin d'obtenir au final une courbe monotone lisse. Elles se présentent sous forme de fonctions cubiques (encadré 2).

Encadré 3. Seuils d'alerte appliqués à l'absence de maladies chez les poulets de chair.

Pour les poulets de chair, les signes cliniques suivants sont évalués en ferme ou en abattoir : ascite, déshydratation, septicémie, hépatite, péricardite, abcès sous-cutané et mortalité. La prévalence de chaque signe clinique est comparée à un seuil d'alerte, défini comme le taux de prévalence au-delà duquel un plan sanitaire est requis au niveau de l'exploitation.

Signe clinique	Seuil d'alerte (% d'animaux atteints)
Ascite	1
Déshydratation	1
Septicémie	1,5
Hépatite	1,5
Péricardite	1,5
Abcès sous-cutané	1
Mortalité (tenant compte du % d'animaux réformés) :	
- Mortalité si réformés < 20%	6
- Mortalité si réformés compris entre 20 et 50%	7
- Mortalité si réformés ≥ 50%	8

Lorsque la prévalence au sein d'une ferme dépasse la valeur correspondant à la moitié du seuil d'alerte, un avertissement est émis. Le nombre d'alertes et d'avertissements relevés sur une exploitation est pris en compte. Il permet de calculer une somme pondérée (avec des poids de 0,33 pour les avertissements et 1 pour les alertes), ensuite convertie en un score au moyen de fonctions I-splines (cf. exemple fourni dans l'encadré 2).

Lorsqu'un critère se subdivise en sous-critères très différents que les chercheurs peuvent difficilement considérer ensemble au premier coup d'œil, des groupes de mesures sont examinés, puis agrégés au moyen d'intégrales de Choquet (cf. section suivante).

Par ailleurs, même si cela est peu fréquent, certaines mesures peuvent se rapporter simultanément à plusieurs critères (par exemple, la maigreur de l'animal peut être associée à une mauvaise alimentation (donc la faim) ou à une maladie (par exemple parasitaire), voire aux deux à la fois). Afin d'éviter de compter deux fois les mêmes problèmes, les mesures ont été affectées à un seul critère, à l'exception de rares cas où l'interprétation qui en est faite varie selon les critères. Ainsi, l'accès des bovins au pâturage permet de vérifier à la fois la facilité de déplacement (notamment pour les animaux attachés en hiver) et l'expression des comportements.

4.2 / Construction des principes à partir des critères

Les critères pris en compte pour un même principe sont agrégés afin de calculer les scores au niveau des principes. Par exemple, les scores obtenus par une exploitation pour les critères «absence de blessures», «absence de maladies» et «absence de douleurs causées par les pratiques d'élevage» sont combinés pour exprimer le degré de conformité de cette

exploitation au principe «bonne santé». La consultation des chercheurs en sciences animales et sociales montre qu'ils considèrent certains critères comme plus importants que d'autres. Ainsi, pour la plupart des espèces animales, la hiérarchie par ordre d'importance entre critères du principe «bonne santé» est la suivante : «absence de maladies», «absence de blessures», «absence de douleurs causées par les pratiques d'élevage». De plus, toutes les personnes consultées sauf une limitaient fortement les compensations entre les scores (par exemple, l'absence de maladies ne compense pas la présence de blessures, et inversement). Un opérateur spécifique, l'intégrale de Choquet, a été employé pour tenir compte de ces deux types de raisonnement concomitants : poids intrinsèques des critères et limitation des compensations entre critères. En bref, l'intégrale de Choquet généralise la notion de moyenne pondérée avec des poids qui peuvent être associés à chaque critère pris séparément, et également à tout ensemble de critères (Grabisch et Roubens 2000). Les scores obtenus pour chacun des critères d'un même principe sont rangés par ordre croissant. L'écart entre le score minimal et le score minimal suivant est calculé et un poids (appelé «capacité») est affecté à cet écart. Cette opération est répétée jusqu'au score le plus élevé. Un exemple de calcul des scores au niveau des principes est fourni dans l'encadré 4. Les capacités ont été calculées de sorte à minimiser le carré des écarts entre le score issu de l'intégrale de Choquet et ceux donnés par les

personnes interrogées. Un exemple de résultats est donné dans l'encadré 5.

4.3 / Agrégation des principes en une évaluation globale

Les scores obtenus par une exploitation pour les quatre principes de bien-être permettent de classer cette exploitation dans une catégorie rendant compte du niveau de bien-être de ses animaux. Le nombre de catégories et leur contenu peuvent varier selon les usages de l'évaluation. Différents usages ont été identifiés après consultation du comité consultatif de Welfare Quality®. Ces usages et leurs implications au regard des différentes catégories de bien-être sont résumés dans le tableau 4. Afin de produire un outil d'évaluation utilisable dans le cadre de ces différents usages, il convient de distinguer quatre catégories de bien-être :

Excellent : le niveau de bien-être des animaux est optimal.

Amélioré (de «Enhanced» dans la version anglaise du protocole) : le bien-être des animaux est bon.

Acceptable : le bien-être des animaux dépasse les exigences minimales requises.

Non classé : le bien-être des animaux est faible et considéré comme inacceptable.

Nous avons décidé d'utiliser une méthode de surclassement afin d'assigner les exploitations à ces catégories

Tableau 4. Usages potentiels de l'évaluation du bien-être.

	Usages potentiels de l'évaluation du bien-être	Catégories de bien-être applicables	
		Nombre	Description
Scénario 1	Critère de conditionnalité des aides et future définition d'un standard minimal du bien-être animal	2	En deçà/au-dessus des exigences légales minimales ou équivalent (en l'absence de législation)
	Étiquetage obligatoire délimitant différents niveaux de bien-être	4	Faible / Normal / Bon / Excellent
Scénario 2	Définition d'objectifs de bien-être dans le cadre de programmes de certification d'exploitations - Étiquetage volontaire	3	Très bon niveau / Haut niveau / Autre (c'est-à-dire insuffisant pour participer à un programme de certification)
Scénario 3	Diagnostic-conseil aux éleveurs permettant le contrôle des résultats de stratégies d'amélioration du bien-être	Plusieurs	Très faible ----- Très bon niveau de bien-être (avec catégories intermédiaires)
Scénario 4	Outil d'auto-évaluation	Plusieurs	Très faible ----- Très bon niveau de bien-être (avec catégories intermédiaires)
	Évaluation de nouveaux systèmes d'élevage		
	Contribution à la recherche dans le domaine du bien-être animal		

Encadré 4. Utilisation de l'intégrale de Choquet dans le calcul du score au niveau du principe «comportement approprié».

Le principe «comportement approprié» se compose de quatre critères : «*expression des comportements sociaux*», «*expression des autres comportements*», «*bonne relation Homme-Animal*» et «*état émotionnel positif*». Dans un premier temps, les scores obtenus par une exploitation sur ces quatre critères sont classés par ordre croissant, du plus mauvais au meilleur. Le calcul du score de principe consiste en l'addition de plusieurs termes. Le premier terme correspond au plus mauvais score. Le second terme correspond à la différence entre le plus mauvais score et le suivant, multipliée par la «*capacité*» du groupe de critères restants, c'est-à-dire tous les critères à l'exception de celui ayant obtenu le plus mauvais score. Le troisième terme correspond à la différence entre le second plus mauvais score et le suivant, multipliée par la «*capacité*» du groupe formé par la combinaison des critères restants, c'est-à-dire tous sauf les deux qui ont obtenu les plus mauvais scores. Le quatrième terme (ici le dernier car on agrège quatre critères) correspond à la différence entre les troisième et quatrième plus mauvais scores (ici le dernier et donc le meilleur), multipliée par la capacité du groupe de critères restants (ici uniquement le dernier). De manière générale, l'opération est répétée jusqu'à ce que le critère ayant obtenu le meilleur score ait été pris en compte. Dans l'exemple du principe «comportement approprié» (composé de quatre critères), la formule comprend quatre termes et se décline en 24 alternatives :

$$\text{Score de principe} = \left\{ \begin{array}{ll} S_a + (S_b - S_a)\mu_{bcd} + (S_c - S_b)\mu_{cd} + (S_d - S_c)\mu_d & \text{si } \overbrace{S_a \leq S_b \leq S_c \leq S_d}^{\text{Classement des critères en fonction de leurs scores}} \\ S_a + (S_b - S_a)\mu_{bcd} + (S_d - S_b)\mu_{cd} + (S_c - S_d)\mu_c & \text{si } S_a \leq S_b \leq S_d \leq S_c \\ S_a + (S_c - S_a)\mu_{bcd} + (S_b - S_c)\mu_{cd} + (S_d - S_b)\mu_d & \text{si } S_a \leq S_c \leq S_b \leq S_d \\ S_a + (S_c - S_a)\mu_{bcd} + (S_d - S_c)\mu_{cd} + (S_b - S_d)\mu_b & \text{si } S_a \leq S_c \leq S_d \leq S_b \\ S_a + (S_d - S_a)\mu_{bcd} + (S_b - S_d)\mu_{cd} + (S_c - S_b)\mu_c & \text{si } S_a \leq S_d \leq S_b \leq S_c \\ S_a + (S_d - S_a)\mu_{bcd} + (S_c - S_d)\mu_{cd} + (S_b - S_c)\mu_b & \text{si } S_a \leq S_d \leq S_c \leq S_b \\ \\ S_b + (S_a - S_b)\mu_{acd} + (S_c - S_a)\mu_{cd} + (S_d - S_c)\mu_d & \text{si } S_b \leq S_a \leq S_c \leq S_d \\ S_b + (S_a - S_b)\mu_{acd} + (S_d - S_a)\mu_{cd} + (S_c - S_d)\mu_c & \text{si } S_b \leq S_a \leq S_d \leq S_c \\ S_b + (S_c - S_b)\mu_{acd} + (S_b - S_c)\mu_{ad} + (S_d - S_b)\mu_d & \text{si } S_b \leq S_c \leq S_a \leq S_d \\ S_b + (S_c - S_b)\mu_{acd} + (S_d - S_c)\mu_{ad} + (S_b - S_d)\mu_a & \text{si } S_b \leq S_c \leq S_d \leq S_a \\ S_b + (S_d - S_b)\mu_{acd} + (S_b - S_d)\mu_{ac} + (S_c - S_b)\mu_c & \text{si } S_b \leq S_d \leq S_a \leq S_c \\ S_b + (S_d - S_b)\mu_{acd} + (S_c - S_d)\mu_{ac} + (S_b - S_c)\mu_a & \text{si } S_b \leq S_d \leq S_c \leq S_a \\ \\ S_c + (S_a - S_c)\mu_{abd} + (S_b - S_a)\mu_{bd} + (S_d - S_b)\mu_d & \text{si } S_c \leq S_a \leq S_b \leq S_d \\ S_c + (S_a - S_c)\mu_{abd} + (S_d - S_a)\mu_{bd} + (S_b - S_d)\mu_b & \text{si } S_c \leq S_a \leq S_d \leq S_b \\ S_c + (S_b - S_c)\mu_{abd} + (S_a - S_b)\mu_{ad} + (S_d - S_a)\mu_d & \text{si } S_c \leq S_b \leq S_a \leq S_d \\ \text{etc.} & \\ \\ S_d + (S_a - S_d)\mu_{abc} + (S_b - S_a)\mu_{bc} + (S_c - S_b)\mu_c & \text{si } S_d \leq S_a \leq S_b \leq S_c \\ S_d + (S_a - S_d)\mu_{abc} + (S_c - S_a)\mu_{bc} + (S_b - S_c)\mu_b & \text{si } S_d \leq S_a \leq S_c \leq S_b \\ S_d + (S_b - S_d)\mu_{abc} + (S_a - S_b)\mu_{ac} + (S_c - S_b)\mu_c & \text{si } S_d \leq S_b \leq S_a \leq S_c \\ \text{etc.} & \end{array} \right.$$

NB : S_a , S_b , S_c et S_d représentent les scores obtenus par une ferme donnée respectivement pour les critères a («*expression des comportements sociaux*»), b («*expression des autres comportements*»), c («*bonne relation Homme-Animal*») et d («*état émotionnel positif*»)

μ_a , μ_b , μ_c , μ_d constituent les capacités respectives pour les critères a , b , c et d

μ_{ab} correspond à la capacité du groupe formé par les critères a et b

μ_{abc} correspond à la capacité du groupe formé par les critères a , b et c

etc.

de bien-être. Avec une méthode de sur-classement, des profils sont définis pour chaque catégorie. Ils sont composés de valeurs de référence sur chaque dimension d'évaluation (ici les quatre principes de bien-être). Puis les alternatives (ici les exploitations) sont comparées à ces profils à l'aide de règles d'appartenance qui déterminent les conditions pour qu'une alternative surclasse un profil (c'est-à-dire peut être considérée

au moins aussi bonne que ce profil) pour être classée dans la catégorie correspondante (Perny 1998).

Dans notre modèle, une valeur de référence est définie pour chaque catégorie de bien-être et chaque principe. Cette valeur représente l'objectif qu'une exploitation doit s'efforcer d'atteindre pour être affectée à cette catégorie. Nous avons d'abord discuté

la possibilité d'avoir des valeurs de référence absolues ou au contraire relatives. Dans le premier cas, les valeurs de références sont prédéfinies, quels que soient les résultats obtenus par les exploitations. Ceci constitue un risque car les valeurs de référence peuvent être fixées à un niveau trop élevé pour être atteintes, ou au contraire, à un niveau trop faible ne permettant pas de discriminer les exploitations entre elles. Dans

Encadré 5. Exemple d'agrégation de scores de critères en score de principe en utilisant l'intégrale de Choquet : cas du principe «comportement approprié» pour les bovins à l'engrais.

Afin d'agrèger les scores obtenus sur les quatre critères composant le principe «comportement approprié» nous avons consulté des personnes en leur présentant un jeu de données virtuelles et en leur demandant d'attribuer un score à chacune des exploitations de ce jeu de données. Leurs avis sont représentés dans le tableau qui suit par le score moyen obtenu par chaque exploitation.

a- Expression des comportements sociaux	b- Expression des autres comportements	c- Bonne relation Homme-Animal	d- État émotionnel positif	Score moyen attribué par les personnes consultées	Score calculé
35	35	65	65	43	43
35	50	50	65	45	45
35	50	65	50	44	44
35	65	35	65	41	41
35	65	50	50	44	43
35	65	65	35	39	39
50	35	50	65	45	45
50	35	65	50	44	45
50	50	35	65	45	45
50	50	50	50	51	50
50	50	65	35	43	42
50	65	35	50	43	43
50	65	50	35	41	42
65	35	35	65	40	40
65	35	50	50	44	44
65	35	65	35	38	38
65	50	35	50	44	44
65	50	50	35	42	42
65	65	35	35	38	38

Sur la base de cette consultation, nous avons déterminé les paramètres de l'intégrale de Choquet (c.-à-d. les valeurs des capacités) permettant de minimiser le carré des écarts entre les scores donnés par les personnes consultées et les scores calculés. Ces paramètres sont présentés dans le tableau ci-dessous :

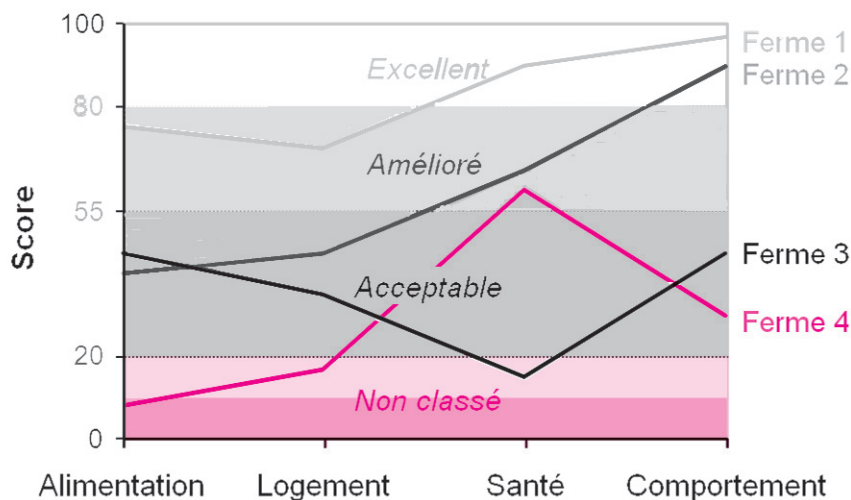
Valeurs des 14 capacités nécessaires au calcul de l'intégrale de Choquet :					
μ_a	μ_b	μ_c	μ_d		
0.06	0.03	0.09	0.15		
μ_{ab}	μ_{ac}	μ_{ad}	μ_{bc}	μ_{bd}	μ_{cd}
0.09	0.09	0.17	0.14	0.18	0.27
μ_{abc}	μ_{abd}	μ_{acd}	μ_{bcd}		
0.41	0.52	0.55	0.51		

Les scores de principes ainsi calculés sont très proches des scores moyens attribués par les personnes consultées, comme on peut le voir en comparant les deux dernières colonnes du premier tableau.

Encadré 6. Agrégation des quatre scores de principes en une évaluation globale de l'exploitation.

Pour passer des quatre principes à l'évaluation globale du bien-être au niveau de l'exploitation, nous avons suivi une méthode dite de «surclassement» reposant sur trois étapes :

1. La définition de plusieurs **catégories de bien-être**. Elle se base sur les utilisations potentielles qui pourraient être faites de l'outil Welfare Quality®. Ainsi, nous avons défini quatre catégories : excellent, amélioré, acceptable et non classé.
2. Le choix de **profils de référence** qui vont délimiter ces quatre catégories. Nous avons opté pour des profils horizontaux fixés à 80 pour la catégorie «Excellent», 55 pour «Amélioré» et 20 pour «Non classé».
3. Le choix de **règles d'appartenance** à chaque catégorie. Dans la figure ci-dessous, quatre exploitations représentatives de chaque catégorie sont représentées, en suivant pour cela le jeu de règles 2/2/3 (tel que défini dans la section 5-3 de l'article).



le second cas, les valeurs de référence correspondent à un certain pourcentage d'exploitations issues d'une population donnée dans laquelle le système est utilisé. Par exemple, on peut décider que les 10% d'exploitations obtenant les meilleurs résultats au sein de cette population, sont excellentes, quel que soit le niveau de bien-être effectivement obtenu. Des valeurs de référence relatives peuvent également être problématiques car : 1) «excellente» peut dans les faits correspondre à un faible niveau de bien-être ; 2) la classification des exploitations dépend de la population observée, puisqu'une exploitation peut être classée «excellente» au sein d'une population et être classée «améliorée» voire «acceptable» dans une autre population ; 3) le suivi dans le temps des progrès réalisés par les exploitations n'est pas possible, par exemple une exploitation peut rester affectée à la catégorie «acceptable», alors que le niveau de bien-être des animaux s'est élevé. Compte tenu de ces éléments, nous avons fait le choix d'utiliser des valeurs de référence absolues, en vérifiant qu'elles étaient réalistes au regard des résultats obtenus par les exploitations visitées dans Welfare Quality®, et qu'elles permettaient bien de discriminer les exploitations entre elles.

Parce que l'échelle de valeur (de 0 à 100) avait la même signification quel que soit le principe de bien-être et qu'il ne semblait pas justifié de considérer certains principes comme plus importants que d'autres, nous avons décidé de définir des profils de référence horizontaux. Autrement dit, une valeur de référence unique délimite une catégorie de bien-être, et s'applique indifféremment aux quatre principes. Ainsi, la valeur de référence associée à la catégorie «acceptable» a été fixée à 20 car cette valeur a un sens bien précis sur l'échelle de 0 à 100, utilisée pour exprimer les scores au niveau des critères et des principes. En effet, une exploitation obtenant un score inférieur à 20 pour l'un des critères ou l'un des principes, a de fortes chances d'être exclue de tout programme de certification sur le bien-être. La valeur 50 possède elle aussi un sens bien défini : une exploitation ayant un score de 50 n'est ni bonne ni mauvaise. C'est pourquoi nous avons placé la valeur de référence délimitant la catégorie «amélioré» à 55. Enfin, la valeur de référence caractérisant la catégorie «excellente» a été déterminée symétriquement à celle de la catégorie «acceptable», soit à 80. Ces profils de référence ont été discutés et validés par le comité directeur et le comité consultatif du projet.

Nous avons ensuite exploré cinq jeux de règles d'appartenance des exploitations aux quatre catégories de bien-être, par comparaison des exploitations aux profils de référence définis ci-dessus. Ces jeux de règles diffèrent par leur sévérité. Une première règle intuitive est «l'unanimité» : une exploitation doit se conformer à la valeur de référence d'une catégorie de bien-être sur tous les principes pour être affectée à cette catégorie. Avec ce jeu de règles, une exploitation est considérée «excellente» si elle obtient un score de 80 ou plus sur tous les principes, elle est considérée «améliorée» si elle obtient un score de 55 ou plus sur tous les principes, elle est considérée «acceptable» si elle obtient un score de 20 ou plus sur tous les principes, sinon elle est «non classée». Ce jeu de règles est appelé 4/4/4, en référence au nombre minimal de principes (ici tous donc 4) pour lesquels l'exploitation doit obtenir des scores au moins égaux aux valeurs de référence. Avec les autres jeux de règles, une exploitation doit obtenir des scores supérieurs à la valeur de référence sur certains principes seulement (1, 2 ou 3 sur 4) et obtenir un score au moins équivalent à la valeur de référence de la catégorie inférieure suivante pour les autres. Les jeux 3/3/3, 2/3/4, 2/2/3 et 1/2/3 ont été

testés. Avec le jeu de règle le plus souple (1/2/3), une exploitation est «*excellente*» si tous ses scores sont supérieurs à 55 et au moins un est égal ou supérieur à 80, une exploitation est «*améliorée*» si tous ses scores sont supérieurs à 20 et au moins deux sont égaux ou supérieurs à 55, elle est «*acceptable*» si tous ses scores sont supérieurs à 10 et au moins trois sont égaux ou supérieurs à 20. Le jeu de règles d'appartenance a été choisi suite aux analyses des jeux de données et à la consultation du comité directeur de Welfare Quality® et de son comité consultatif (voir section 5.3).

5 / Tests sur les jeux de données

Dans Welfare Quality®, pour chaque type d'animaux, plusieurs dizaines d'exploitations ont été visitées (91 fermes laitières, 85 élevages de bovins allaitants, 224 exploitations de veaux de boucherie, 90 naisseurs porcins, 71 engraisseurs porcins, 77 élevages de poules pondeuses et 58 élevages de poulets de chair). Pour les vaches laitières, les partenaires du projet nous ont fourni des jeux de données complets (c'est-à-dire comprenant toutes les mesures nécessaires au calcul de l'évaluation globale) sur 69 exploitations de trois pays différents (20 en Allemagne, 24 en Autriche et 25 en Italie).

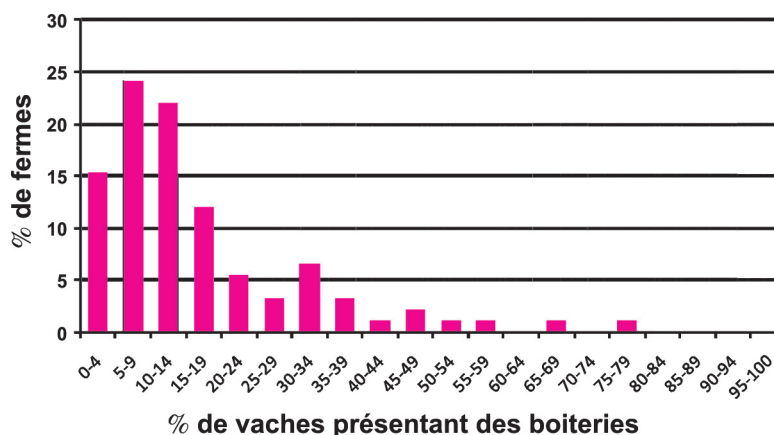
5.1 / Des mesures aux critères

Pour définir les scores associés aux différents critères pour les vaches laitières, nous avons organisé deux cycles de consultation : un avant la collecte des données sur le terrain et un après. Lors de la confrontation des évaluations initiales des chercheurs consultés avec la répartition des résultats des exploitations laitières, il est apparu que les chercheurs avaient porté un jugement relativement sévère. Par exemple, 50% des fermes avaient plus de 12% des vaches boitant (ce % étant pondéré en fonction de la sévérité des boiteries) et leur score sur ce critère s'établissait à 30 ou en-deçà. Nous avons présenté la distribution des boiteries aux chercheurs consultés (figure 2), en leur demandant s'ils souhaitaient modifier leurs évaluations. La plupart se sont montrés moins sévères durant ce second cycle de consultation (tableau 5). Néanmoins, ils ont considéré les boiteries comme un problème grave pour les vaches, et selon leur avis, 50% des fermes obtiennent un score inférieur à 44. Nous avons modifié les paramètres du modèle en nous appuyant sur la seconde consultation.

En ce qui concerne les bovins à l'engrais, les veaux de boucherie, les porcs à

Figure 2. Distribution des boiteries dans les exploitations laitières participant au projet Welfare Quality®

Le pourcentage de vaches présentant des boiteries est pondéré par gravité (poids de 0,29 pour les boiteries légères et 1 pour les boiteries sérieuses).



l'engrais et les poulets de chair, le calcul des scores a été obtenu à partir d'une seule consultation car celle-ci s'est déroulée après la collecte des données en fermes.

5.2 / Des critères aux principes

Pour définir les scores associés aux différents principes pour les vaches laitières, nous avons également réalisé deux cycles de consultation : un avant et un après la collecte des données dans les exploitations laitières. Les évaluations se sont révélées quasi identiques. Nous avons par conséquent décidé de nous limiter à une seule consultation pour les autres types d'animaux, sans confronter les scores des chercheurs interrogés aux résultats effectivement obtenus par les exploitations.

5.3 / Agrégation des principes en une évaluation globale

Nous avons analysé la distribution des scores obtenus par les exploitations laitières sur les quatre principes. Nous avons constaté qu'aucune ferme laitière n'était en mesure d'atteindre le score de 80 pour tous les principes (correspondant à l'excellence). En fait, ce score

n'a été enregistré que sur le principe «bonne alimentation» et sur 18% des fermes observées. En revanche, une large variété de résultats, de 0 à 70, a été relevée sur les autres principes. Nous avons par conséquent choisi de conserver les trois valeurs de référence définies (20, 55 et 80), tout en maintenant l'idée que si le système d'évaluation est mis en place, les progrès devront être encouragés et des scores équivalents ou supérieurs à 80 obtenus.

Nous avons ensuite testé les cinq jeux de règles d'appartenance destinées à déterminer le niveau de performance suffisant pour permettre à une exploitation d'être affectée à une catégorie donnée de bien-être. Le jeu 4/4/4 ne semble guère réaliste puisqu'avec un tel jeu de règles, la moitié des fermes visitées dans Welfare Quality® se retrouve «non classée» tandis que l'autre moitié est seulement «acceptable» (voir résultats obtenus pour la règle 4/4/4 dans la figure 3). Les règles qui reflètent au mieux l'évaluation générale des observateurs ayant visité les fermes, permettent de discriminer les exploitations et étant en harmonie avec les attentes du comité consultatif sont les suivantes (jeu de règles 2/2/3) :

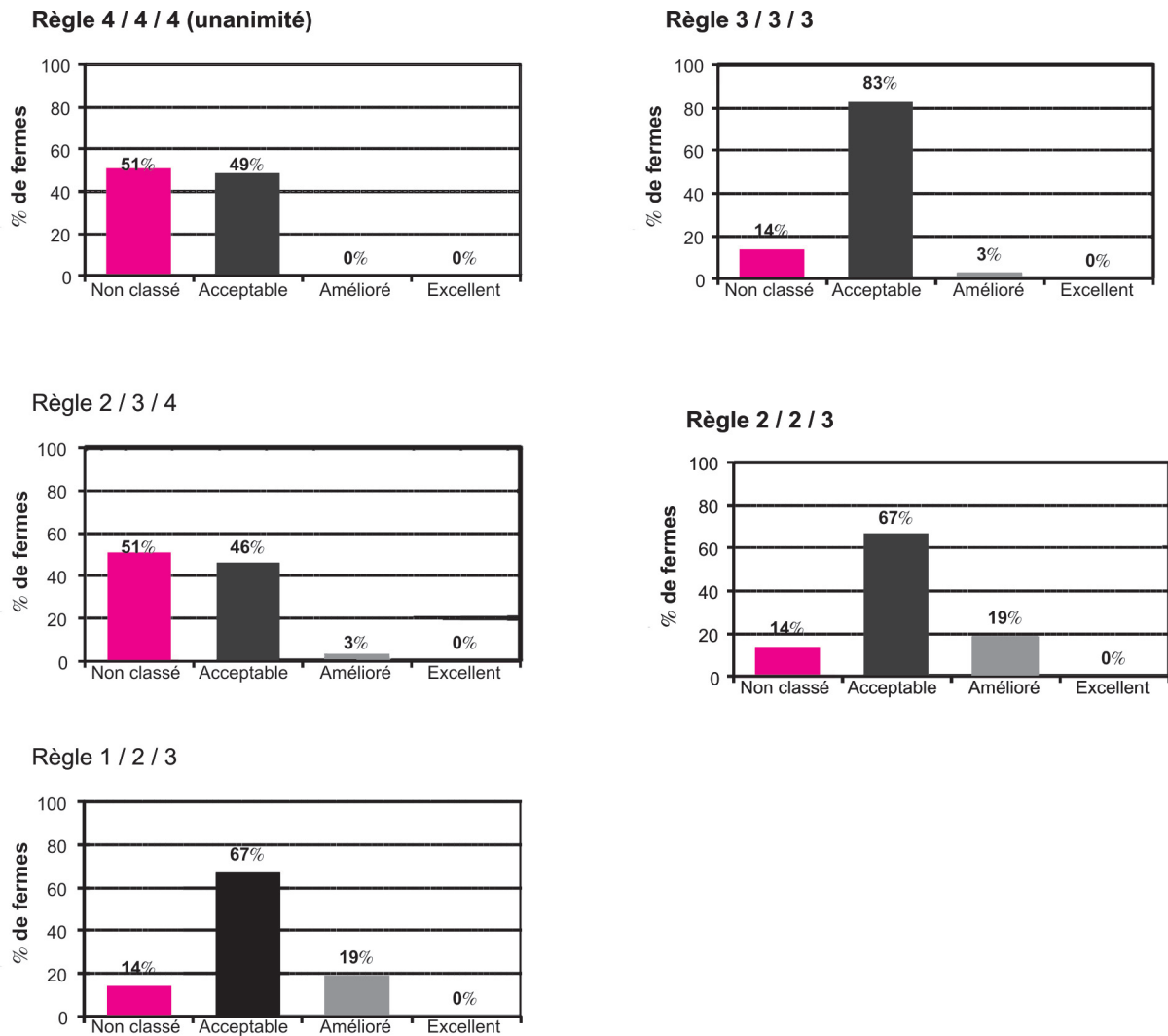
Tableau 5. Comparaison des scores attribués par les experts sur les boiteries avant et après connaissance de leur distribution parmi les fermes laitières visitées.

% de vaches présentant des boiteries ¹	Score moyen des experts lors de la 1 ^{ère} consultation	Score moyen des experts lors de la 2 ^{nde} consultation
10	34	46
30	13	20
60	6	6

¹ Pondéré par gravité (poids de 0,29 pour les boiteries légères et 1 pour les boiteries sérieuses).

Figure 3. Répartition de 69 fermes laitières européennes parmi les quatre catégories de bien-être, en fonction de différentes règles.

Pour chaque règle, les chiffres indiquent dans quelle mesure une exploitation doit se conformer à la valeur de référence délimitant une catégorie pour y être affectée. Les jeux de règles d'appartenance 4/4/4 (unanimité), 3/3/3, 2/3/4, 2/2/3 (jeu de règles finalement retenu pour le modèle d'évaluation) et 1/2/3 sont décrites dans la section 5-3 de l'article.



Une exploitation est affectée à la catégorie «*excellent*» lorsque son score est supérieur à 55 sur l'ensemble des principes et dépasse 80 sur deux d'entre eux. Elle est rangée dans la catégorie «*amélioré*» si son score est supérieur à 20 sur l'ensemble des principes et dépasse 55 sur deux d'entre eux. Les exploitations entrant dans la catégorie «*acceptable*» sont celles qui enregistrent un score supérieur à 10 sur tous les principes et dépassent 20 sur trois d'entre eux. Les exploitations n'obtenant pas ces notes minimales ne sont pas classées (encadré 6).

Ces règles ont été retenues dans le modèle d'évaluation. Toutefois, un seuil d'indifférence séparant les écarts significatifs de ceux qui ne le sont pas et équivalent à 5 s'applique : par exemple, 50 n'est pas considéré comme significativement distinct de 55. La valeur de ce seuil a été décidée d'après les variations entre les personnes consultées pour

construire les critères. L'erreur standard moyenne entre les scores attribués par ces personnes aux mêmes jeux de données est de 5,13, d'où le 5 du seuil d'indifférence.

6 / Retour aux questions éthiques

Lors des consultations, nous n'abordions pas directement les choix éthiques. Le raisonnement éthique des personnes consultées était exprimé au travers de leurs réactions vis-à-vis des jeux de données présentés. Le modèle d'évaluation a été soumis à une analyse éthique *a posteriori* avec le concours de P. Sandøe, philosophe de l'université de Copenhague. En analysant les résultats produits par le modèle, nous avons identifié à quels choix éthiques ils correspondaient.

Tout d'abord, force est de constater que le modèle met l'accent sur les animaux en moins bon état. Ceci se retrouve dans le calcul des scores au niveau des critères. Ainsi, comme le montre la figure de l'encadré 2, une proportion de 9% de vaches présentant une boiterie sévère au sein d'un élevage aboutit à un score de 50. Or, si la même importance était accordée aux animaux en bon et en mauvais état, la même ferme obtiendrait un score de 50 avec 50% de vaches boitant sévèrement. Cependant, le bien-être moyen du groupe est aussi pris en compte. Ainsi, une ferme dont 5% des vaches boitent sévèrement et 50% boitent légèrement obtient un score de boiterie inférieur à une ferme où 10% de vaches boitent sévèrement et aucune ne boite légèrement. Les mêmes résultats sont constatés quels que soient les critères et les espèces animales. Il apparaît, par

conséquent, que le modèle d'évaluation résulte d'un compromis entre la priorité accordée aux animaux en moins bon état et le niveau moyen de bien-être du troupeau. Ce raisonnement est d'ailleurs conforme à celui qui est souvent adopté pour arbitrer des décisions concernant la protection des humains (Arneson 2008).

Comme nous l'avons vu précédemment, les compensations entre critères demeurent très limitées. Elles ne sont cependant pas impossibles. Une exploitation ayant obtenu un score inférieur à 20 sur un critère a ainsi une chance de figurer dans la catégorie «acceptable» en atteignant un score élevé sur d'autres critères. Par exemple, un élevage noté 42 pour l'alimentation (60 pour l'absence de faim et 40 pour l'absence de soif), 39 pour le logement (50 pour le repos, 25 pour le confort thermique, 75 pour la facilité de déplacement), 20 pour la santé (14 pour l'absence de blessures, 38 pour l'absence de maladies et 52 pour l'absence de douleurs), et 43 pour le comportement (35 pour les comportements sociaux et autres, 65 pour la relation Homme-Animal et l'état émotionnel positif) serait considéré comme acceptable en n'ayant obtenu qu'un score de 14 sur le critère «absence de blessures». Après consultation interne au projet et des jurys de citoyens-consommateurs, ceci ne semble guère en phase avec l'idée générale que la plupart des personnes se font du bien-être, selon laquelle tous les critères doivent produire un score d'un certain niveau minimal (Miele *et al* accepté). Pour se rapprocher de cette opinion largement partagée par les scientifiques comme par les citoyens, il faudrait ajouter une règle supplémentaire : une exploitation ne pourrait être considérée comme acceptable si son score était inférieur à 20 sur l'un des critères pris en compte. Cependant, l'application d'une telle règle semble prématurée, puisqu'elle entraînerait l'exclusion de la moitié des exploitations.

Au final, le modèle d'évaluation est le fruit, d'une part, d'attentes théoriques sur ce qui doit être considéré comme un haut vs bas niveau de bien-être animal, et d'autre part des résultats que l'on peut raisonnablement escompter dans les exploitations à l'heure actuelle. Les attentes théoriques sont reflétées au travers des valeurs de référence délimitant les catégories de bien-être (20 pour «acceptable», 55 pour «amélioré» et 80 pour «excellent»). La réalité du terrain a orienté le choix des règles pour classer les exploitations dans les différentes catégories de bien-être (par exemple, une exploitation doit avoir un score excellent sur au moins

deux principes et au minimum amélioré sur les autres pour être notée comme globalement excellente). Le modèle d'évaluation semble donc offrir un équilibre satisfaisant entre aspirations théoriques et pragmatisme. Dans les recommandations finales du projet, nous avons conseillé que la règle d'appartenance à la catégorie «acceptable» soit révisée quand des progrès suffisants en termes de bien-être des animaux auront été réalisés (par exemple seulement 10% des exploitations obtiennent des scores inférieurs à 20), de sorte à ne pas considérer une exploitation acceptable si elle obtient un score inférieur à 20 sur un critère.

Conclusions

Au sein du projet Welfare Quality[®], le travail présenté dans cet article a permis de développer des méthodes standardisées d'évaluation globale du bien-être applicables en exploitations de vaches laitières, jeunes bovins, veaux de boucherie, porcs à l'engrais et poulets de chair. Pour les autres types d'animaux pris en compte dans le projet (poules pondeuses et truies) et pour la période de transport-abattage, les mesures ont été développées, mais la procédure d'évaluation n'a pas encore été finalisée.

La démarche d'évaluation implique de porter un jugement. Pour proposer un modèle d'évaluation globale du bien-être animal, il nous fallait donc définir une stratégie claire de consultation des intervenants sollicités pour porter ce jugement. Au cours du projet Welfare Quality[®], plusieurs groupes de personnes ont été consultés : des chercheurs en sciences animales afin d'interpréter les données recueillies sur les animaux ou leur environnement et afin d'apporter, dans une certaine mesure, le point de vue des animaux ; des chercheurs en sciences sociales afin d'intégrer la perception des porteurs d'enjeux concernés par l'évaluation du bien-être des animaux (consommateurs/citoyens, distributeurs, éleveurs) ; et les porteurs d'enjeux eux-mêmes. Ces derniers ne sont intervenus qu'au dernier stade de construction du modèle d'évaluation afin d'éviter qu'ils ne soient à la fois juges et parties.

L'apport des méthodologies d'évaluation multicritère a été décisif pour rendre compte des différentes appréciations. Nous avons écarté les approches permettant des compensations complètes entre différents aspects du bien-être. Par ailleurs, nous avons eu recours à une méthode de surclassement permet-

tant de discriminer les exploitations selon quatre catégories de bien-être : «non classé», «acceptable», «amélioré» et «excellent», chacune correspondant à un type de décision potentielle des porteurs d'enjeux (rejet/approbation/distinction).

Certes le modèle reflète les avis des personnes consultées. Se pose donc la question de sa validité dans d'autres contextes. Toutefois, les personnes consultées avaient des bagages scientifiques et/ou culturels variés, bien que toutes étaient européennes, et nous avons constaté que les écarts entre leurs avis étaient faibles (l'erreur standard entre les scores de critères attribués par les différentes personnes était de 5,1% en moyenne, et seulement de 2,3% pour l'agrégation des scores de critères en scores de principes), ce qui suggère que le modèle est robuste. Par ailleurs, une fois le modèle d'évaluation élaboré, celui-ci a été discuté auprès de jurys de citoyens et d'éleveurs au Royaume-Uni, en Italie, aux Pays-Bas et en Norvège et les résultats confortent les choix majeurs réalisés lors de la construction du modèle, en particulier la priorité donnée aux animaux en moins bon état et la faible compensation entre critères (Miele *et al* accepté). Les mesures servant à l'évaluation ont également été testées en Amérique du Sud dans des élevages extensifs et la plupart sont transposables dans de telles conditions.

Certains outils utilisés en matière de certification ou de contrôle d'exploitations sont fondés sur des exigences minimales à respecter (par exemple les standards de certification d'EurepGAP). De tels systèmes peuvent reposer sur une logique du «tout ou rien» (échec ou réussite), dans laquelle un éleveur ayant échoué peut difficilement savoir à quel point il est éloigné d'un résultat positif. L'outil d'évaluation Welfare Quality[®] évite cet écueil en proposant quatre catégories distinctes et en encourageant ainsi les exploitants à s'engager dans la voie d'une amélioration continue du niveau de bien-être des animaux.

En jugeant rétrospectivement le travail accompli, nous conseillons aux personnes souhaitant mettre au point de tels modèles de travailler en étroite collaboration avec les chercheurs chargés de définir les indicateurs. Il est en effet essentiel de posséder une idée claire de l'architecture sous-tendant le modèle d'évaluation pour choisir les indicateurs. Il convient également de s'assurer que les indicateurs retenus peuvent être interprétés facilement au regard du bien-être (ou d'un

autre objet que l'on souhaite appréhender).

Les difficultés que nous avons rencontrées pour établir le modèle d'évaluation ne sont certainement pas spécifiques au concept de bien-être animal. D'autres concepts, comme la durabilité des systèmes de production, partagent de nombreuses propriétés avec le concept de bien-être (qu'il inclut d'ailleurs) : multiples critères, indicateurs variés, nécessité de limiter les compensations entre critères,

etc. Le travail que nous avons réalisé pour l'évaluation globale du bien-être des animaux peut aider la conception de modèles d'évaluation d'autres concepts multidimensionnels appliqués à l'agriculture.

Remerciements

Le présent travail fait partie du projet de recherche Welfare Quality® co-financé par la Commission européenne, au sein du 6^{ème} programme cadre, contrat n° FOOD-CT-2004-506508. Ce texte

représente le point de vue des auteurs et pas nécessairement une position de la Commission qui ne peut être tenue responsable de l'utilisation qui pourra être faite de l'information. Les auteurs remercient le groupe de travail chargé du développement du modèle, par ordre alphabétique : M. Bonde, M.B.M. Bracke, A. Butterworth, J. Capdeville, G. De Rosa, B. Engel, L.J. Keeling, T. Rousing, J.T. Sørensen et C.J. van Reenen, pour leur participation active dans le développement du modèle d'évaluation.

Références

- Arneson R., 2008. Egalitarianism, In: The Stanford Encyclopedia of Philosophy. Zalta E.N. (Ed).
- Bekoff M., 2008. Why «good welfare» isn't «good enough»: minding animals and increasing our compassionate footprint. ARBS Ann. Rev. Biomed. Sci., Theme Topic on «Unraveling Animal Welfare», 10, T1-T14.
- Botreau R., Bonde M., Butterworth A., Perny P., Bracke M.B.M., Capdeville J., Veissier I., 2007a. Aggregation of measures to produce an overall assessment of animal welfare. Part 1 - A review of existing methods. *Animal*, 1, 1179-1187.
- Botreau R., Bracke M.B.M., Perny P., Butterworth A., Capdeville J., van Reenen C.G., Veissier I., 2007b. Aggregation of measures to produce an overall assessment of animal welfare: Part 2 - Analysis of constraints. *Animal*, 1, 1188-1197.
- Botreau R., Veissier I., Butterworth A., Bracke M.B.M., Keeling L.J., 2007c. Definition of criteria for overall assessment of animal welfare. *Anim. Welfare*, 16, 225-228.
- Botreau R., Capdeville J., Perny P., Veissier I., 2008. Multicriteria evaluation of animal welfare at farm level; an application to MCDA methodologies. *Foundations of Computing and Decision Sciences*, 33, 287-316.
- Botreau R., Veissier I., Perny P., 2009. Overall assessment of cow welfare: strategy adopted in Welfare Quality®. *Anim. Welfare*, 18, 363-370.
- Bouyssou D., 1990. Building criteria: a prerequisite for MCDA. In: Readings in multiple criteria decision-aid. Bana e Costa C.A. (Ed), Springer Verlag, Heidelberg, 58-80.
- Bouyssou D., Marchant T., Pirlot M., Perny P., Tsoukias A., Vincke P., 2000. Evaluation and decision models a critical perspective. Kluwer's academic publishers, Dordrecht, The Netherlands, 274p.
- Bracke M.B.M., Spruijt B.M., Metz J.H.M., 1999. Overall animal welfare assessment reviewed. Part 1: Is it possible? *Neth. J. Agric. Sci.*, 47, 273-291.
- Broom D.M., Blokhuis W.J., Canali E., Dijkhuizen A.A., Fallom R., Le Neindre P., Saloniemi H., Webster A.J.F., 1995. Report of the scientific veterinary committee. *Animal welfare section on the welfare of calves*. Commission Européenne, 120p.
- Commission Européenne, 2002. Communication de la commission au conseil et au parlement européen sur la législation applicable au bien-être des animaux d'élevage dans les pays tiers et sur son incidence pour l'UE, COM, 626, 49p.
- European Commission, 2005. Attitudes of consumers towards the welfare of farmed animals. Special Eurobarometer 229/ Wave 63.2 - TNS opinion and social, http://ec.europa.eu/public_opinion/archives/ebs/ebs_229_en.pdf, 138p.
- European Commission, 2007a. Attitudes of consumers towards the welfare of farmed animals. Wave 2. Special Eurobarometer 229(2)/Wave 64.4 - TNS Opinion & Social, http://ec.europa.eu/food/animal/welfare/survey/sp_barometer_fa_en.pdf, 60p.
- European Commission, 2007b. Attitudes of EU citizens towards Animal Welfare - Special eurobarometer 270/Wave 66.1 - TNS Opinion & Social, http://ec.europa.eu/food/animal/welfare/survey/sp_barometer_fa_en.pdf, 51p.
- Farm Animal Welfare Council, 1992. FAWC updates the five freedoms. *Vet. Rec.*, 17, 357.
- Fraser D., 1995. Science, values and animal welfare: exploring the «inextricable connection». *Anim. Welfare*, 4, 103-117.
- Fraser D., Weary D.M., Pajor E.A., Milligan B.N., 1997. A scientific conception of animal welfare that reflects ethical concerns. *Anim. Welfare*, 6, 187-205.
- Grabisch M., Roubens M., 2000. Application of the Choquet integral in multicriteria decision making. In: Fuzzy measures and integrals. Theory and applications. Grabish M., Murofushi T., Sugeno M. (Eds), *Studies in Fuzziness and Soft Computing*, Physica-Verlag, Heidelberg, Germany, 348-374.
- Holm L., Ritz C., Ladewig J., 2007. Measuring animal preferences: Shape of double demand curves and the effect of procedure used for varying workloads on their crosspoint. *Appl. Anim. Behav. Sci.*, 107, 133-146.
- Miele M., Veissier I., Evans A., Accepté. *Animal welfare: establishing a dialogue between science and society*. *Anim. Welfare*, special issue *Knowing animals*.
- Napolitano F., Girolami A., Braghieri A., 2010. Consumer liking and willingness to pay high welfare animal-based products. *Trends Food Sci. Technol.* sous presse.
- Perny P., 1998. Multicriteria filtering methods based on concordance and non-discordance principles. *Ann.Operat. Res.*, 80, 137-165.
- Polten R., 2007. Proceedings of the workshop Animal welfare improving by labelling?, Polten R. (Ed). 28 March 2007, Brussels, Belgium, 86p.
- Roy B., 1985. *Méthodologie Multicritère d'Aide à la Décision*. Roy B. (Ed.). Economica, Paris, France, 423p.
- Roy B., 1991. The outranking approach and the foundations of ELECTRE methods. *Theory and Decision*, 31, 49-73.
- Roy B., 1993. Decision Science or Decision-Aid Science? *Eur. J. Operat. Res.*, 66, 184-204.
- Veissier I., Jensen K.K., Botreau R., Sandøe P., 2010. Highlighting ethical choices underlying the scoring of animal welfare in the Welfare Quality® scheme. *Anim. Welfare*, special issue *Knowing animals*.
- Webster A.J.F., 2009. The Virtuous Bicycle: a delivery vehicle for improved farm animal welfare. *Anim. Welfare*, 18, 141-147.
- Welfare Quality®, 2009a. Welfare Quality® assessment protocol for cattle (fattening cattle, dairy cows, veal calves). Welfare Quality® Consortium, Lelystad, The Netherlands, 182p.
- Welfare Quality®, 2009b. Welfare Quality® assessment protocol for pigs (sows and piglets, growing and finishing pigs). Welfare Quality® Consortium, Lelystad, The Netherlands, 114p.
- Welfare Quality®, 2009c. Welfare Quality® assessment protocol for poultry (broilers, laying hens). Welfare Quality® Consortium, Lelystad, The Netherlands, 114p.
- Whay H.R., Main D.C.J., Green L.E., Webster A.J.F., 2003. An Animal-based welfare assessment of group-housed calves on UK dairy farms. *Anim. Welfare*, 12, 611-617.

Résumé

Face au développement en Europe de programmes de certification d'élevages mettant en avant le respect du bien-être animal, le projet Welfare Quality® a proposé à l'Union Européenne d'établir un modèle d'évaluation globale des exploitations (fermes ou abattoirs) au regard du bien-être des animaux. Ce modèle repose sur une évaluation multicritère dans laquelle 12 critères de bien-être, regroupés en 4 principes, ont été définis. Après avoir identifié les mesures nécessaires pour vérifier qu'une exploitation est conforme aux différents critères, le modèle a été progressivement construit. Il aboutit à un classement des exploitations en quatre catégories reflétant le niveau de bien-être (d'excellent à très faible). Pour ce faire, des méthodes multicritères ont été utilisées, évitant ainsi les écueils d'une agrégation naïve de l'information (comme les sommes pondérées qui autorisent les pleines compensations entre critères, ce qui n'est pas compatible avec le concept de bien-être). Différentes personnes (chercheurs en sciences animales ou sociales, porteurs d'enjeux) ont été consultées au moyen de jeux de données virtuelles sur lesquelles elles devaient donner une appréciation. Le modèle a été développé et paramétré afin de refléter ces appréciations. Une analyse *a posteriori* des résultats produits montre que le modèle accorde la priorité aux animaux en moins bon état tout en tenant compte de l'état moyen du troupeau, et qu'il limite fortement les compensations entre critères. Cet exemple montre qu'en suivant cette méthodologie rigoureuse, il est possible d'aider l'évaluation d'un concept multidimensionnel, même si cet exercice comporte des choix de valeurs qu'il convient d'explicitier et de modéliser. La méthodologie suivie pour construire ce modèle d'évaluation du bien-être peut être transposée pour évaluer d'autres concepts multidimensionnels comme la durabilité des systèmes de production.

Abstract

Multicriteria evaluation applied to farm animal welfare: difficulties and solutions from the Welfare Quality® project

More and more often European schemes for farm certification imply the fulfillment of animal welfare. The Welfare Quality® project aimed at proposing an overall assessment model of animal welfare at the unit level (farm or slaughterhouse). This model relies on a multicriterion evaluation. First, 12 welfare criteria, grouped into 4 principles, were defined. After identifying the measures needed to check that an animal unit complies with the various criteria, the model was progressively constructed so as to finally classify animal units into four categories reflecting the level of animal welfare (from excellent to very poor). We used multicriteria evaluation methods, trying to avoid the pitfalls of a naïve aggregation of the information (e.g. weighted sums authorizing full compensations between criteria, not compatible with the concept of welfare). Different persons (animal and social scientists, stakeholders) were consulted. They were presented virtual datasets to which they had to assign scores. The model was developed and parameterized so as to reflect their opinions. A post-hoc analysis of the results showed that the model gives the priority to animals in bad condition while also taking into account the average state of the group of animals, and that compensation between criteria are strongly limited. Following this rigorous methodology we were able to propose an evaluation of a multidimensional concept. Such an exercise requires value judgments that must be clarified and modelled. The methodology followed to construct this animal welfare assessment model could be transposed to the assessment of other multidimensional concepts like the sustainability of production systems.

VEISSIER I., BOTREAU R., PERNY P., 2010. Evaluation multicritère appliquée au bien-être des animaux en ferme ou à l'abattoir : difficultés et solutions du projet Welfare Quality®. *Inra Prod. Anim.*, 23, 269-284.

