



HAL
open science

Mesure de l'appréciation des aliments par les consommateurs : état des pratiques et propositions méthodologiques

Isabelle Boutrolle

► **To cite this version:**

Isabelle Boutrolle. Mesure de l'appréciation des aliments par les consommateurs : état des pratiques et propositions méthodologiques. Sciences du Vivant [q-bio]. AgroParisTech, 2007. Français. NNT : 2007AGPT0083 . pastel-00004525

HAL Id: pastel-00004525

<https://pastel.hal.science/pastel-00004525>

Submitted on 13 Jan 2009

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

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

AgroParisTech (centre de Massy)

Ecole doctorale ABIES

Département Science de l'Aliment

THESE

Présentée pour obtenir le grade de Docteur

Discipline : SCIENCES ALIMENTAIRES

par

Isabelle BOUTROLLE

MESURE DE L'APPRECIATION DES ALIMENTS PAR LES CONSOMMATEURS

Etat des pratiques et propositions méthodologiques

Directeur de thèse : Marc Danzart

Co-directeur de thèse : Julien Delarue

Thèse soutenue publiquement le 11 mai 2007

Membres du jury

M. Herbert Meiselman
Mme Sylvie Issanchou
M. Thierry Meyer
M. Bernard Ruffieux
Mme Delphine Arranz
M. Marc Danzart

Président
Rapporteur
Rapporteur
Examineur
Examineur
Directeur de thèse

REMERCIEMENTS

Même si l'accomplissement d'une thèse est généralement associé à un travail solitaire, ce projet aura été pour moi une expérience humaine particulièrement riche de collaborations étroites avec un grand nombre de personnes auxquelles je souhaite témoigner toute ma gratitude.

Mes premiers remerciements vont à **Julien Delarue** pour sa grande disponibilité, son exigence et son soutien permanent. *« Merci pour ton extrême motivation, ton implication et ton engouement pour ce travail qui m'ont donné confiance et m'ont encouragé à dépasser certaines barrières qui me semblaient insurmontables dans ce parcours de thèse. En tant que « première thésarde », je tiens ici à témoigner de ton encadrement exemplaire dont j'ai eu la chance de bénéficier. »*

Je tiens également à remercier **Delphine Arranz** pour son intérêt tout au long de ce travail et pour son acharnement à débloquer du budget pour chacune de mes expérimentations. *« Merci d'avoir continué à suivre ce travail jusqu'à son aboutissement alors que tes préoccupations professionnelles t'emmenaient vers d'autres contrées. Merci également pour le climat de confiance que tu as installé entre nous et pour tes précieux conseils tant professionnels que personnels. »*

Toute ma gratitude va également à **Marc Danzart** pour son encadrement et son ouverture. *« Merci pour les propositions de traitements statistiques que tu as imaginé chaque fois que je venais solliciter ton aide avec un nouveau jeu de données. »*

Je remercie très sincèrement **Ep Köster** et **Pierre Combris** pour l'intérêt qu'ils ont immédiatement porté à cette thèse. *« Merci pour votre disponibilité et votre regard avisé d'expert lors du développement des protocoles ainsi que pour votre aide précieuse concernant la communication de ces travaux. »*

Merci à **Sylvie Issanchou**, **Thierry Meyer**, **Bernard Ruffieux** et **Herbert Meiselman** pour m'avoir fait l'honneur d'accepter de lire, critiquer et évaluer mon manuscrit. *« Chacune de vos remarques soulève de futurs travaux de recherche tous plus captivants les uns que les autres »*

Un énorme merci également à **Romain Monrozier** et **Michel Rogeaux** qui m'ont permis de dégager et clarifier les pistes de recherche les plus pertinentes sans perdre de vue

l'application industrielle de ce travail. *« Merci pour votre confiance et le temps que vous avez consacré à ce travail. »*

Merci à toute l'équipe **Consumer Science** de Danone pour leur accueil et l'ambiance de travail chaleureuse. Merci plus particulièrement à monsieur Alain Montembault pour la liberté qu'il m'a laissé, notamment dans l'organisation de mon temps de travail.

Je tiens également à remercier chaleureusement **Yves Keromnes** de la société Sky Consulting qui a coordonné la mise en place d'une grande partie des expérimentations de ce travail. *« Merci d'avoir eu la délicatesse de ne jamais émettre de doute quant à la possibilité d'appliquer mes protocoles farfelus. Ta confiance et ton ouverture ont été précieuses dans ce travail. »*

Merci à **Gaëlle, Caroline et Thomas**, étudiants de l'ENSIA, qui ont mené avec un profond enthousiasme la mise en place des tests en situation de petit déjeuner et en situation sociale.

Ma gratitude ne saurait bien évidemment passer à côté de **ma famille** et de **mes proches** dont le soutien et l'encouragement quotidien m'ont permis de ne jamais baisser les bras. Merci tout particulièrement à **Sébastien** d'avoir supporté mes sautes d'humeur ainsi que ma passivité dans notre vie quotidienne durant les long mois de rédaction.

Enfin, plus que des remerciements, je tiens à exprimer ma profonde amitié à mes co-thésards, **Bérengère** et **Pierrick**. *« Merci Bérengère pour ton aide sur l'élaboration des questionnaires. Merci Pierrick d'avoir supporté mes questions incessantes sur le traitement des données. Enfin, merci à vous deux pour ces trois années de tendre complicité pour lesquelles je serai presque prête (je dis bien presque) à refaire une thèse. »*

SOMMAIRE

<i>Avant – propos</i>	8
-----------------------	---

CHAPITRE I : ANALYSE DU PROBLEME	12
---	-----------

A - LES TESTS ORGANOLEPTIQUES : CONTEXTE INDUSTRIEL

I - ENJEUX STRATEGIQUES : INTERET DE LA MESURE HEDONIQUE POUR LES ENTREPRISES AGROALIMENTAIRES	13
II - ETAT DES PRATIQUES POUR LA COLLECTE DE DONNEES HEDONIQUES	17
II.1 - <i>Quelle mesure hédonique ?</i>	17
II.1.1 - <i>Les tests de préférence</i>	17
II.1.2 - <i>Les tests de notation</i>	18
II.1.3 - <i>Les questions diagnostiques</i>	19
II.2 - <i>Quel groupe de sujets ?</i>	20
II.2.1 - <i>Qui interroger ?</i>	20
II.2.2 - <i>Combien de sujets ?</i>	21
II.3 - <i>Quelles conditions de test ?</i>	22
II.3.1 - <i>Les tests en salle et en laboratoire</i>	23
II.3.2 - <i>Les tests à domicile</i>	26

B - IMPLICATION DES VARIABLES DE CONTEXTE DANS LA FORMATION DU JUGEMENT HEDONIQUE

I - LE JUGEMENT EVALUATIF : PROCESSUS INTEGRATIF A LA BASE DE L'ATTITUDE	29
I.1 - <i>L'attitude : définition et processus de construction</i>	29
I.2 - <i>Processus de construction du jugement hédonique d'un aliment</i>	30
II - L'INFLUENCE DU CONTEXTE DANS CE PROCESSUS INTEGRATIF	33
II.1 - <i>Définition</i>	33
II.2 - <i>Les effets de contexte dans les enquêtes attitudinales</i>	34
II.2.1 - <i>Par quels mécanismes le contexte influence-t-il le jugement ?</i>	34
II.2.2 - <i>Comment les effets de contexte se manifestent-ils ?</i>	36
II.3 - <i>Typologie des variables de contexte dans le cadre de la perception hédonique d'un aliment</i>	37
II.3.1 - <i>Typologie des variables de contexte selon leur source</i>	37
II.3.2 - <i>Typologie des variables de contexte selon leur caractéristique « interne » ou « externe »</i>	38
II.3.3 - <i>Typologie des variables de contexte selon leur caractéristique temporelle</i>	39
II.3.4 - <i>Vers une globalisation des variables de contexte</i>	39

C - MISE EN EVIDENCE DE L'INFLUENCE DU CONTEXTE SUR LA MESURE HEDONIQUE

I - INFLUENCE DE L'ENVIRONNEMENT GLOBAL DE CONSOMMATION SUR LA MESURE HEDONIQUE	42
I.1 - <i>Une mesure hédonique sensible au lieu de consommation</i>	42
I.1.1 - <i>Impact systématique du lieu de consommation sur la mesure hédonique</i>	42
I.1.2 - <i>Comment le lieu de consommation influence-t-il la mesure hédonique ?</i>	43
I.2 - <i>Une mesure hédonique sensible au caractère artificiel des conditions de dégustation</i>	47
I.2.1 - <i>Les conditions de dégustation artificielles n'aboutissent pas systématiquement à des résultats différents de ceux obtenus dans des conditions naturelles</i>	47
I.2.2 - <i>Nature de l'impact des conditions de dégustation sur les résultats</i>	48
I.2.3 - <i>Est-il possible d'anticiper les conséquences de la mise en place d'un test en conditions artificielles ou naturelles de dégustation sur la mesure hédonique ?</i>	49

II - IDENTIFICATION DES VARIABLES DE CONTEXTE SUSCEPTIBLES D'INFLUENCER LA MESURE HEDONIQUE	50
II.1 - Influence des facteurs contextuels externes	50
II.1.1 - La quantité d'aliment consommée lors de la dégustation	51
II.1.2 - La préparation de l'aliment testé	51
II.1.3 - L'environnement alimentaire	53
II.1.4 - L'information disponible sur l'aliment testé	53
II.1.5 - Le moment de consommation	54
II.1.6 - L'environnement social	55
II.1.7 - La possibilité de choisir l'aliment dégusté	56
II.1.8 - L'effort physique nécessaire pour obtenir l'aliment	57
II.1.9 - Le contexte évaluatif apporté par les autres produits testés	59
II.1.10 - Le design du questionnaire	60
II.2 - Influence des facteurs contextuels internes	
II.2.1 - Le vécu alimentaire de l'individu	65
II.2.2 - L'état interne de l'individu	67
III - BILAN : DISCUSSION SUR LA PERTINENCE DES REPONSES HEDONIQUES COLLECTEES EN CONDITIONS STANDARDISEES D'EVALUATION	70
D - IMPACT DES CONDITIONS STANDARDISEES D'EVALUATION SUR LA MESURE HEDONIQUE : CAS DES PRODUITS DANONE	
<hr/>	
I - TEST EN SALLE VS. TEST A DOMICILE : DES RESULTATS DIFFERENTS EN FONCTION DE LA CATEGORIE DE PRODUITS	74
ARTICLE 1 : "Central Location Test vs. Home Use Test: contrasting results depending on product type"	
II - TEST EN SALLE VS. TEST A DOMICILE : CAS D'UNE PROBLEMATIQUE DE REDUCTION DE SUCRE	97
II.1 - Matériels et méthodes	97
II.1.1 - Les produits	97
II.1.2 - Les sujets	97
II.1.3 - Les protocoles	98
II.1.4 - Analyse des données	99
II.2 - Résultats	99
II.2.1 - Influence du contexte sur l'appréciation globale des produits	99
II.2.2 - Influence du contexte sur la notation de l'intensité sucrée par rapport à l'idéal	100
II.3 - Discussion	101
III - TEST EN SALLE VS. TEST A DOMICILE : DES RESULTATS DIFFERENTS EN FONCTION DU SEXE DES SUJETS	103
III.1 - Matériels et méthodes	103
III.1.1 - Les produits	103
III.1.2 - Les sujets	103
III.1.3 - Les protocoles	104
III.1.4 - Analyse des données	104
III.2 - Résultats	105
III.2.1 - Analyse des données de l'ensemble du panel	105
III.2.2 - Analyse des données générées par les hommes	106
III.2.3 - Analyse des données générées par les femmes	107
III.3 - Discussion	107
IV - TEST EN SALLE VS. TEST A DOMICILE : DISCUSSION GENERALE	111
IV.1 - Impact des conditions d'évaluation sur le niveau de notation des produits	111
IV.1.1 - Hypothèses interprétatives	111
IV.1.2 - Implications concernant l'interprétation de la modification de l'appréciation	114
IV.2 - Impact des conditions d'évaluation sur les conclusions statistiques des tests	116
IV.2.1 - Instabilité de l'influence du contexte selon les études	116
IV.2.2 - Stabilité de l'influence du contexte selon la procédure de présentation des produits	119
IV.2.3 - Description de l'influence du contexte sur les conclusions statistiques des tests	119

A - PROBLEMATIQUE INDUSTRIELLE

I - TEST EN SALLE vs. TEST A DOMICILE : « CONFLICTING DESIDERATA »	122
I.1 - Test en salle	122
I.2 - Test à domicile	122
I.3 - Conflicting desiderata	124
II - OBJECTIF	126
III - CHOIX DES AXES METHODOLOGIQUES EXPLORES	127
III.1 – Les approches orientées « situation »	128
III.1.1 - Des conditions physiques de consommation plus appropriées	128
III.1.2 - Une réponse hédonique mesurée selon différentes situations de consommation	128
III.2 – Les approches orientées « implication »	129
III.2.1 - Le test d'authenticité	131
III.2.2 - Les méthodes non-hypothétiques	131

B - LES APPROCHES ORIENTEES « SITUATION »

I - DES CONDITIONS DE CONSOMMATION PLUS APPROPRIEES : LES TESTS SITUATIONNELS	132
ARTICLE 2 : “Some attempts to improve realism in Central Location Tests”	
II - UNE REPOSE HEDONIQUE MESUREE SELON DIFFERENTES SITUATIONS DE CONSOMMATION EVOQUEES : UTILISATION DE LA MEMOIRE EPISODIQUE DU PARTICIPANT	162
II.1 - Contexte et objectif	162
II.1.1 - Diversité des conditions de consommation et de leur signification	162
II.1.2 - Aspect multi-situationnel de la consommation des produits	164
II.1.3 - Démarche générale	167
II.2 - Matériels et méthodes	167
II.2.1 - Les produits	167
II.2.2 - Les sujets	168
II.2.3 - La procédure	168
II.2.4 - Analyse des données	172
II.3 - Résultats et discussion	173
II.3.1 - Considérations méthodologiques	173
II.3.2 - Etape 1 : appréciation globale après amorçage autobiographique	174
II.3.3 - Etape 2 : intention de consommation selon différentes situations évoquées	176
II.3.4 - Etape 1 vs. étape 2 : appréciation globale vs. intention de consommation	182
II.4 - Discussion générale	186

C - LES APPROCHES ORIENTEES « IMPLICATION »

I - LE TEST D’AUTHENTICITÉ	190
I.1 - Principe de la méthode d'authenticité	190
I.2 - Application de la méthode pour comparer les performances organoleptiques de deux biscuits feuilletés au fromage	191
ARTICLE 3 : “How to improve respondents’ commitment when evaluating food acceptance: use of the authenticity test”	
II - LES METHODE NON-HYPOTHETIQUES	213
II.1 - Principe et description des méthodes non-hypothétiques	213
II.1.1 - Principe général	213
II.1.2 - Les différents types de méthodes non-hypothétiques	215
II.2 - Choix du protocole non-hypothétique à privilégier	220
II.3 - Application du protocole pour comparer les performances organoleptiques de deux biscuits feuilletés au fromage	222
ARTICLE 4 : “Proposition d’amélioration des protocoles de test de produits alimentaires”	

A - QUESTIONS METHODOLOGIQUES

I - NOTE HEDONIQUE ET/OU PREFERENCE ?	253
II - EFFECTIF DU PANEL	254
III - PRESENTATION DES PRODUITS : MONADIQUE PUR OU MONADIQUE SEQUENTIEL ?	256
IV - LES QUESTIONS DIAGNOSTIQUES	258
IV.1 - Des informations hédoniques supplémentaires	258
IV.2 - Comment introduire ces questions dans les protocoles ?	259
IV.3 - Quelles questions diagnostiques ?	260
V - EVALUATION PONCTUELLE ET DYNAMIQUE DES PREFERENCES	260
VI - QUESTIONS METHODOLOGIQUES RELATIVES AU TEST EN SALLE	262
VI.1 - La préparation de l'aliment testé	262
VI.2 - Etat physiologique des participants et quantité consommée	263

B - IMPACT DES CONDITIONS ARTIFICIELLES DE DEGUSTATION SUR LA MESURE HEDONIQUE

I - DES NOTES HEDONIQUES EN SALLE MOINS ELEVEES QU' A DOMICILE	264
II - DIFFICILE D' ANTICIPER L' IMPACT DES CONDITIONS ARTIFICIELLES DE DEGUSTATION SUR LES CONCLUSIONS STATISTIQUES DES TESTS	266
II.1 - Les conclusions statistiques ne sont pas systématiquement différentes de celles obtenues lors d'un test à domicile	266
II.2 - Variabilité de la nature de l'influence des conditions d'évaluation sur les conclusions statistiques des tests	267

C - DISCUSSION SUR LES CRITERES DE QUALITE D'UNE METHODE DE TEST CONSOMMATEURS

I - LE CRITERE D' EXACTITUDE DE LA METHODE	268
I.1 - Discussion sur l'utilisation du test à domicile comme méthode de référence	268
I.2 - Proposition de méthodes de référence basées sur des données comportementales	270
II - LE CRITERE DE REPETABILITE DES SUJETS	272
III - LE POUVOIR DISCRIMINANT DE LA METHODE	273
IV - LE CRITERE DE ROBUSTESSE	274

D - PROPOSITIONS METHODOLOGIQUES : APPLICATION EN ROUTINE ?

I - LES APPROCHES ORIENTEES « SITUATION »	275
I.1 - Les tests situationnels : des méthodologies à simplifier	275
I.2 - Protocole à privilégier en fonction de l'univers de consommation du produit	276
I.3 - Proposition de différents protocoles basés sur l'utilisation de la mémoire épisodique	277
II - LES APPROCHES ORIENTEES « IMPLICATION »	279
II.1 - Quel est le test qui permet l'implication la plus forte des participants ?	279
II.2 - Protocole orienté « implication » à privilégier selon le cas d'étude	282
II.2.1 - Le test d'authenticité	282
II.2.2 - Le test de choix expérimentaux non-hypothétiques	283
II.2.3 - Synthèse	284

CONCLUSION GENERALE	285
REFERENCES BIBLIOGRAPHIQUES	290
ANNEXES	

AVANT - PROPOS

AVANT-PROPOS

Ce travail de thèse concernant les pratiques de la mesure de l'appréciation des aliments par les consommateurs a été mené dans le cadre d'une convention CIFRE en partenariat avec le service « *Consumer Science* » du centre de recherche de Danone et le « *Laboratoire de Perception Sensorielle et Sensométrie* » de l'ENSIA (devenu depuis AgroParisTech). Cette collaboration a été par ailleurs particulièrement enrichie par l'intervention et le suivi régulier du professeur Ep Köster, du « *Center for Innovative Consumer Studies* » de l'université de Wageningen (Pays-Bas) et de Pierre Combris, directeur de recherche du « *Laboratoire de Recherche sur la Consommation* » de l'INRA.

Les entreprises agroalimentaires soucieuses d'améliorer leur position concurrentielle sur le marché ont besoin d'accéder aux performances organoleptiques des recettes qu'ils développent et commercialisent. En effet, maximiser l'appréciation d'un aliment basée sur ses caractéristiques sensorielles permet d'augmenter la probabilité d'un ré-achat du produit par le consommateur et in fine d'assurer son succès financier. Ainsi, les tests consommateurs focalisés sur les propriétés sensorielles des aliments sont utilisés sur l'ensemble de la vie d'un produit, de sa phase de développement avant son lancement jusqu'à sa maintenance une fois qu'il est lancé sur le marché. Cependant, nous assistons aujourd'hui à une forte remise en question du pouvoir prédictif des données hédoniques telles qu'elles sont généralement récoltées c'est-à-dire par la génération d'une opinion après une consommation ponctuelle de l'aliment dans des conditions de dégustation standardisées et artificielles. C'est pourquoi il est plutôt recommandé dans la littérature d'utiliser des protocoles de test qui proposent un contact prolongé entre le consommateur et l'aliment dans des conditions naturelles de consommation. Néanmoins, ces méthodologies de test sont compliquées et coûteuses à mettre en place dans un contexte industriel routinier et leur utilisation reste donc exceptionnelle.

Ce travail propose une réflexion méthodologique basée sur les tests consommateurs pratiqués en entreprise et plus particulièrement chez Danone. La démarche globale de ce travail de recherche s'articule autour de deux grands axes (figure 1).

- La première étape vise à éclaircir, par le biais d'un examen de la littérature et d'expérimentations, les risques réels que prennent les industriels en choisissant de mettre en place des tests consommateurs en conditions artificielles (test en salle) plutôt qu'en situations naturelles de consommation (test à domicile).

- La deuxième partie de ce travail concerne une phase plus exploratoire dans laquelle nous proposons des protocoles originaux pour récolter, par le biais d'un test en conditions contrôlées, des données hédoniques représentatives d'un jugement émis lors d'un acte naturel de consommation.

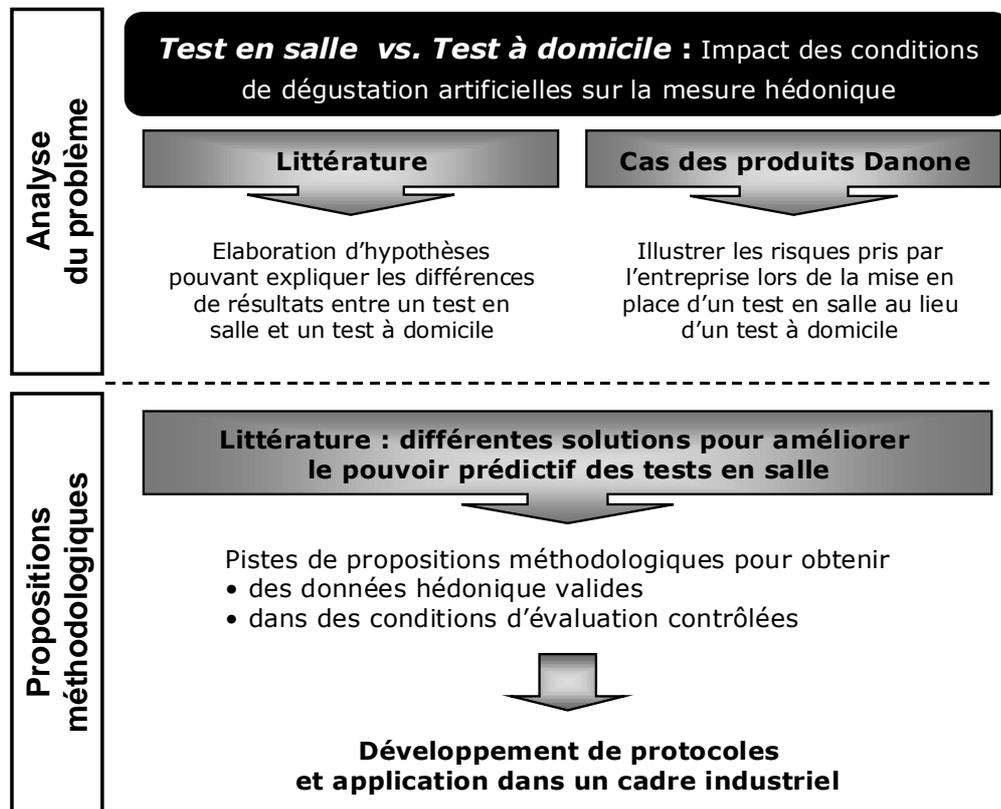


Figure 1 : Démarche globale du travail de recherche

Ce manuscrit est structuré en trois parties et s'articule autour de quatre publications. Certaines parties expérimentales du travail n'ont pas été rédigées sous forme d'article et sont également présentées dans ce document.

- Le chapitre I présente la mise en évidence et l'analyse du problème de validité externe des données hédoniques collectées lors d'un test en conditions artificielles de dégustation.
- Le chapitre II présente la nécessité de développer des protocoles de test alternatifs puis détaille les applications des propositions méthodologiques explorées.
- Le chapitre III synthétise les questions méthodologiques et fondamentales soulevées tout au long de ce travail et propose des éléments de réponse.

PUBLICATIONS

Article 1

Boutrolle I., Delarue J., Arranz D., Rogeaux M., Köster E.P. (2007). Central Location Test vs. Home Use Test: contrasting results depending on product type. *Food Quality and Preference*, vol 18, p. 490-499.

Article 2

Boutrolle I., Delarue J., Arranz D., Rogeaux M., Köster E.P., Danzart, M. Some attempts to improve realism in Central Location Tests. *Food Quality and Preference*. Article en préparation.

Article 3

Boutrolle I., Delarue J., Köster E.P., Arranz D., Danzart M. How to improve respondents' commitment when evaluating food acceptance: use of the authenticity test. *Food Quality and Preference*. Soumis en octobre 2006.

Article 4

Boutrolle I., Combris P., Abiven F., Arranz D., Danzart M., Delarue J. Proposition d'amélioration des protocoles de test de produits alimentaires. *Revue Française de Marketing*. Accepté en octobre 2007.

Article 5 (annexe I)

Boutrolle I., Arranz D., Rogeaux M., Delarue J. (2005). Comparing central location test and home use test results: application of a new criterion. *Food Quality and Preference*, vol 16, p. 704-713.

PARTICIPATIONS AUX CONGRES**Communications orales**

Boutrolle I., Arranz D., Rogeaux M., Delarue J. (2004). Comparer les résultats d'un test en salle et d'un test à domicile : élaboration d'un nouveau critère. In *8^{ème} journées agro-industrie et méthodes statistiques (Agrostat)*, 10-12 mars, Rennes, France.

Boutrolle I., Delarue J., Arranz D., Combris P., Danzart M. (2006). Assessing preference for a low-involvement food product using a real-choice experiment. Comparison with stated preference. In *IAREP/SABE Conference "Behavioural Economics and Economic Psychology"*, 5-8 juillet, Paris, France.

Boutrolle I., Delarue J., Arranz D., Köster E. P., Danzart M. (2006). How to improve respondents' commitment when evaluating food acceptance: use of the authenticity test. In *A Sense of Diversity - 2nd European Conference on Sensory Consumer Science of Food and Beverages*, 26-29 septembre, The Hague, NL.

Posters (annexes II et III)

Boutrolle I., Arranz D., Rogeaux M., Danzart M., Delarue J. (2005). Central Location Test vs. Home Use Test: contrasting results depending on product type. In *6th Pangborn Sensory Science Symposium*, 7-11 août, Harrogate, UK.

Boutrolle I., Delarue J., Arranz D., Monrozier R., Combris P., Danzart M. (2006). Development of an incentive-compatible choice experiment to measure low value food product liking. In *International Conference on Food Choice*, 19-22 avril, Birmingham, UK.

Chapitre I

ANALYSE DU PROBLEME

Approche bibliographique et expérimentale

A - LES TESTS ORGANOLEPTIQUES : CONTEXTE INDUSTRIEL**I - ENJEUX STRATEGIQUES : INTERET DE LA MESURE HEDONIQUE POUR LES ENTREPRISES AGROALIMENTAIRES**

Le but de toute entreprise est de satisfaire son client qui sera alors source de profit pour l'entreprise. En effet, ce client satisfait sera susceptible de racheter le produit et de générer un bouche à oreille positif (Anderson, 1998). D'après Rust *et al.* (2002), il est financièrement plus intéressant pour une entreprise d'opter pour une stratégie axée sur la satisfaction des clients plutôt que de chercher à réduire les coûts. Ainsi, face à la dynamique du marché des biens de grande consommation et donc aux contraintes de la concurrence, les exigences du consommateur ont pris une place importante dans le choix et l'élaboration des stratégies d'entreprise. Apparu au milieu des années 60 aux Etats Unis, le mouvement consumériste est probablement le principal catalyseur de cet engouement pour la satisfaction des consommateurs. Il est défini par Kotler (1974) comme « un mouvement organisé de citoyens et de gouvernements qui vise à renforcer les droits et le pouvoir des acheteurs en relation avec des vendeurs ». Il a eu pour conséquence la reconnaissance du droit à l'information et à la protection des consommateurs. Ainsi, né de l'insatisfaction des consommateurs, ce mouvement a permis de rappeler que la satisfaction du client constitue le but ultime de toute organisation commerciale.

Dans ce même objectif de satisfaction des consommateurs comme garant de ré-achat et donc de profit, les entreprises agroalimentaires dépensent une partie de leur budget pour évaluer l'opinion des consommateurs envers les produits qu'elles commercialisent.

La figure 2 récapitule les différentes étapes qui interviennent dans une décision d'achat lorsque le consommateur n'est pas confronté physiquement à l'aliment avant de faire son choix, ce qui est le cas d'une grande partie des situations d'achat. Nous pouvons voir présentés en amont du processus d'achat deux types d'informations déterminantes dans le comportement d'achat d'un produit alimentaire : les données intrinsèques à l'aliment (caractéristiques organoleptiques) et les données extrinsèques à l'aliment.

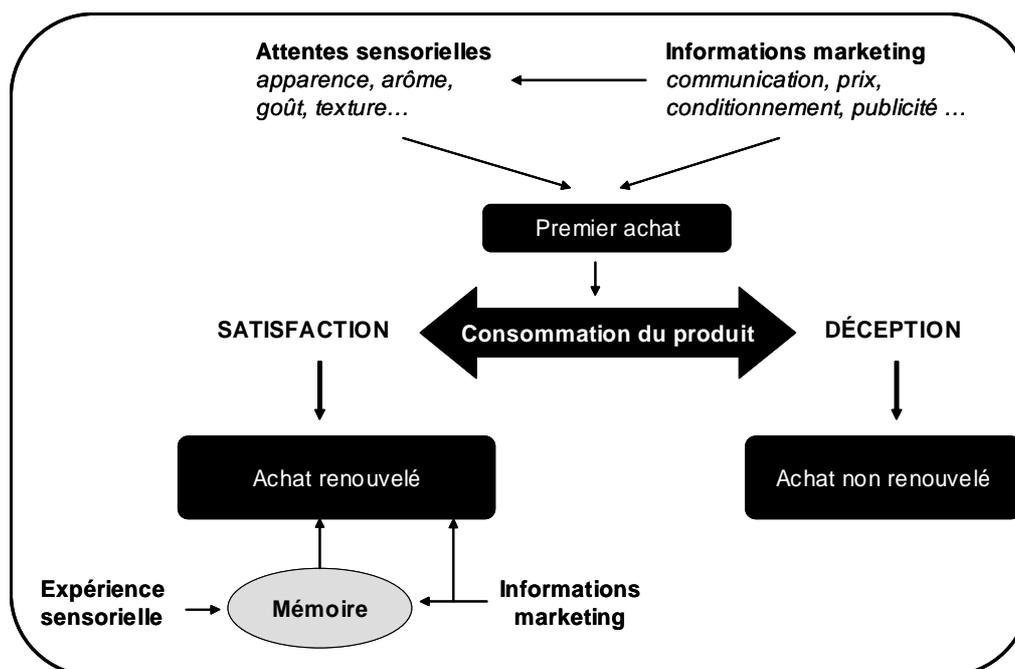


Figure 2 : Etapes intervenant dans une décision d'achat d'un aliment

Lors d'un premier achat, les propriétés organoleptiques de l'aliment interviennent uniquement par le biais des attentes sensorielles générées par la vision du produit dans son emballage. Les données extrinsèques sont donc prépondérantes et à la base de ce premier achat dans lequel le rôle du marketing passe alors par la mise en place de campagnes de communication et de prix attractifs. En revanche, une fois que le consommateur a expérimenté l'aliment, les caractéristiques organoleptiques du produit entrent en jeu de manière décisive dans le comportement de ré-achat par l'intermédiaire de la mémorisation de l'expérience sensorielle du produit et donc de son appréciation (Buck, 2003 ; Grunert, 2003). En effet, dès les années 1950, Young (1948) supportait l'idée que le choix alimentaire est en grande partie déterminé par les propriétés sensorielles des aliments. De la même façon, Chiva (1992) décrit la prise alimentaire comme une activité quotidienne parée de différentes vertus (survie, maintien de la santé...) et de plaisir en précisant que « manger est un acte agréable faisant disparaître les sensations désagréables provoquées par la faim ». L'aliment semble donc particulièrement indissociable de la notion de plaisir. Les propriétés intrinsèques de l'aliment (c'est-à-dire ses caractéristiques sensorielles) et son appréciation sont donc présentes dans l'ensemble des modèles représentant les déterminants du comportement alimentaire (Randall et Sanjur, 1981 ; Shepherd, 1985) dont la figure 3 présente un exemple. La maximisation de l'appréciation de l'aliment par le consommateur (attitude envers les propriétés sensorielles) est donc indispensable pour assurer la survie d'un produit alimentaire

sur le marché. C'est pourquoi, par l'intermédiaire des tests organoleptiques, les entreprises agroalimentaires souhaitent accéder à une information hédonique liée uniquement à l'appréciation de la recette en présentant anonymement les produits débarrassés de toute information extrinsèque (économique, conceptuelle, liée au bénéfice, etc).

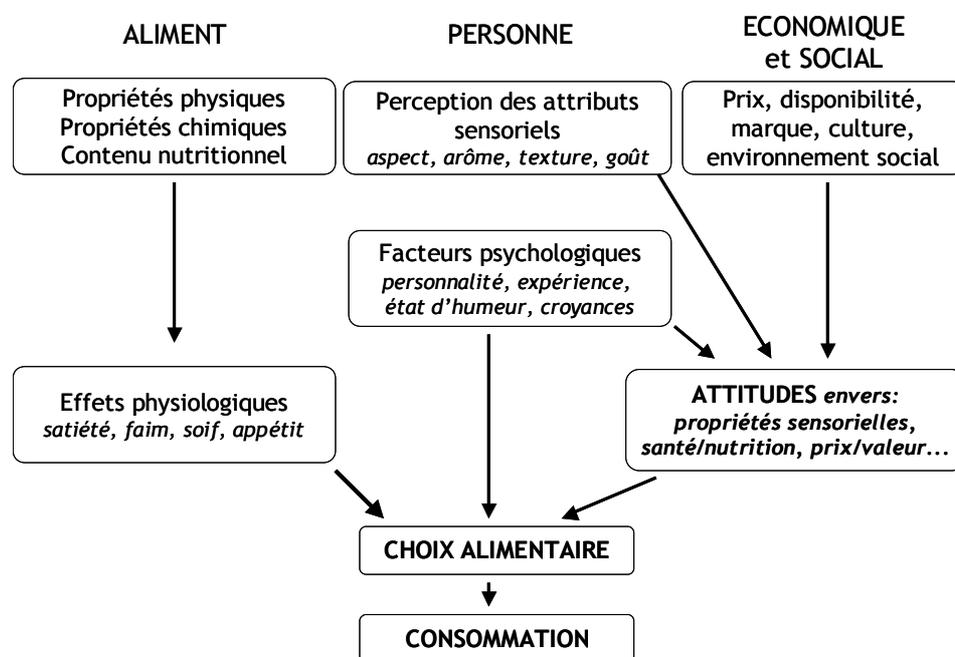


Figure 3 : Facteurs influençant le choix alimentaire (d'après Shepherd, 1985)

Le rôle non négligeable de l'appréciation d'un produit sur son ré-achat entraîne les industriels à extrapoler les performances organoleptiques des recettes obtenues en tests organoleptiques et à les utiliser pour prédire la performance future du produit sur le marché. Cependant, le processus de ré-achat du produit reste également dépendant d'autres informations extrinsèques à la recette de l'aliment. En effet, bien que la communauté scientifique sensorielle affirme que le plaisir sensoriel a un impact sur le comportement d'achat ou sur le choix, l'importance de ce rôle reste à déterminer en comparaison avec celui des autres composants du comportement alimentaire. D'après Fantino (1992), la composante hédonique est le moteur essentiel du comportement alimentaire. Cabanac (1995), de son côté, montre que la maximisation du plaisir alimentaire peut également passer par le souhait de dépenser moins d'argent. L'exemple du succès des produits *hard discount* vient confirmer cet aspect. De manière plus radicale, Buck (2003) restreint l'utilisation des mesures hédoniques obtenues par les tests organoleptiques en précisant que ces tests ne doivent pas être mis en place pour prédire le comportement des consommateurs sur les points de vente mais uniquement pour tester l'impact de changements de formulations, de process ou de

conditionnement sur l'appréciation de la recette. Avec un discours plus tempéré et globalement soutenu par la communauté sensorielle, Lawless et Heymann (1999) précisent qu'une recette non performante en test organoleptique empêchera probablement le succès du produit en dépit d'une campagne marketing rondement menée mais qu'un produit ayant obtenu des scores hédoniques élevés ne sera en revanche pas automatiquement un succès financier.

Quoi qu'il en soit, assurer la supériorité et suivre les performances organoleptiques d'un produit alimentaire reste important pour l'entreprise qui le commercialise. L'utilité et l'importance accordée aux informations hédoniques récoltées lors des tests hédoniques en aveugle tout au long de la vie d'un produit justifient donc ce travail de thèse concernant l'amélioration de la prédiction de ces données hédoniques collectées par les professionnels de l'analyse sensorielle. Nous allons détailler dans les paragraphes suivants les pratiques courantes concernant la mise en place de ces tests hédoniques en aveugle par les industries agroalimentaires.

II - ETAT DES PRATIQUES POUR LA COLLECTE DE DONNEES HEDONIQUES

II.1 - QUELLE MESURE HEDONIQUE ?

De manière théorique deux types de mesures existent pour accéder à l'opinion des consommateurs. Il s'agit des mesures basées sur du questionnement (mesures déclaratives ou attitudinales) ou des mesures basées sur de l'observation (mesures comportementales). Les méthodes d'observation ont l'avantage que les comportements sont observés dans des situations naturelles de consommation et qu'il n'y a normalement pas d'interaction entre l'observateur et le consommateur. Outre des méthodes complexes basées sur des enregistrements vidéo, les mesures comportementales peuvent également être plus simples comme la mesure de la quantité consommée. Bien que très populaires dans le monde universitaire, ces méthodes comportementales sont peu exploitées par les utilisateurs industriels des tests hédoniques compte tenu de leur difficulté de mise en place. Nous nous intéresserons dans ce paragraphe uniquement aux mesures déclaratives et à la façon dont elles sont obtenues en routine pour aider les développeurs lors de problématiques de développement de nouveaux produits, d'amélioration de produit, de réduction de coût ou encore de positionnement par rapport à la concurrence. Ces tests sont mis en place pour comparer la performance organoleptique globale de différentes recettes. Pour cela, deux grands types de méthodes sont disponibles : les tests de préférence et les tests de notation. Par ailleurs, il arrive également que les commanditaires de tests organoleptiques souhaitent récolter des informations concernant certaines caractéristiques sensorielles spécifiques des produits par l'intermédiaire de questions diagnostiques. Nous allons décrire ces trois types de mesures les plus fréquemment utilisées pour accéder à l'appréciation des aliments par les consommateurs.

II.1.1 - Les tests de préférence

Comme son nom l'indique le test de préférence a pour objectif de déterminer un classement de préférence entre les produits dégustés. Les tests de préférence peuvent se décliner sous deux formes : les tests de préférence par paire ou les tests de classement. Le test de préférence par paire consiste à présenter uniquement deux produits en même temps au sujet qui doit indiquer le produit qu'il préfère. Cette méthode peut cependant être utilisée lors d'une étude comparant plus de deux produits (test par paires multiples). Le test de classement,

quant à lui, consiste à présenter directement l'ensemble des produits au sujet qui doit donner un classement de ces produits selon son appréciation.

L'avantage des tests de préférence concerne principalement leur facilité de réalisation par les interviewés. Ils sont donc particulièrement appropriés dans le cas d'études réalisées avec des personnes ayant des difficultés à comprendre les instructions de test (personnes qui ne savent pas lire, enfants) ou encore dans le cas d'études à l'échelle internationale de manière à ne pas obtenir des données biaisées par des systèmes de notation différents. Cependant, les données hédoniques qui découlent des tests de préférence possèdent de nombreuses limites. En effet, ces méthodes ne donnent aucune indication sur le niveau d'appréciation des produits. Une recette peut être préférée à une autre sans pour autant être à un niveau d'appréciation acceptable. Par ailleurs, ces données de classement ne renseignent pas non plus sur l'écart d'appréciation entre les produits. Ainsi, les tests de notation sont souvent privilégiés par rapport aux tests de préférence.

II.1.2 - Les tests de notation

Les tests de notation permettent de déterminer le statut hédonique d'un produit en demandant aux sujets de « noter » les produits dégustés. Ces tests peuvent donc être réalisés sur un seul produit et ne nécessitent pas la comparaison par rapport à d'autres produits. Ces données récoltées dans l'absolu sont particulièrement avantageuses pour les instituts d'études qui peuvent proposer à leurs clients la création de bases de données et ainsi s'assurer une collaboration sur le long terme.

Les tests de notation impliquent l'utilisation d'une échelle de notation. Différents types d'échelles ont été proposés pour collecter le niveau d'appréciation des produits. L'échelle peut être structurée ou non structurée, numérique, sémantique ou picturale. La seule règle indispensable au bon traitement des données « quantitatives » est que l'échelle proposée doit être une échelle d'intervalle c'est-à-dire satisfaisant la règle de l'égalité des intervalles délimités par les différents barreaux de l'échelle. La plupart des procédures statistiques telles que le calcul des moyennes et des variances, les analyses de variances, les régressions et corrélations peuvent alors être utilisées pour analyser les données. La littérature offre de nombreux exemples d'échelles mais l'échelle hédonique à 9 points (Jones *et al.*, 1955 ; Peryam et Pilgrim, 1957) reste la plus utilisée dans la littérature (malgré les doutes émis sur son caractère quantitatif). La figure 4 présente l'équivalent en français et proposé par la norme XP V 09-500 (AFNOR, 2000) en cours de révision.

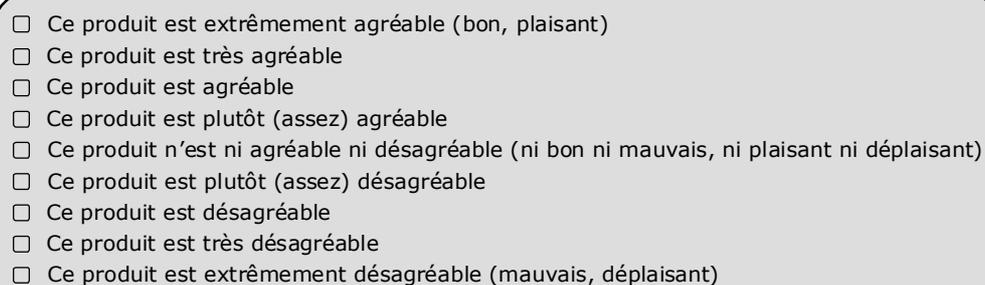
- 
- Ce produit est extrêmement agréable (bon, plaisant)
 - Ce produit est très agréable
 - Ce produit est agréable
 - Ce produit est plutôt (assez) agréable
 - Ce produit n'est ni agréable ni désagréable (ni bon ni mauvais, ni plaisant ni déplaisant)
 - Ce produit est plutôt (assez) désagréable
 - Ce produit est désagréable
 - Ce produit est très désagréable
 - Ce produit est extrêmement désagréable (mauvais, déplaisant)

Figure 4 : Echelle hédonique à 9 points en français (AFNOR, 2000)

Cependant, la pratique des études consommateurs réalisées par l'intermédiaire des instituts d'étude ainsi qu'une analyse récente des échelles utilisées dans les enquêtes de satisfaction ne montre pas une utilisation très fréquente de cette échelle (Boutrolle *et al.*, 2003). Les Français étant habitués dès leur enfance à un système de notation, de nombreuses entreprises françaises ont fait le choix de demander l'opinion aux consommateurs par l'intermédiaire d'une note sur 10 ou sur 20.

II.1.3 - Les questions diagnostiques

Il est aujourd'hui admis que l'objectif des tests hédoniques est d'obtenir l'opinion des consommateurs concernant l'appréciation globale des produits testés. Ainsi, il peut sembler logique de limiter le questionnaire à une seule question portant sur l'appréciation globale ou sur la préférence. Cependant, en pratique, les commanditaires des études hédoniques s'intéressent également à l'opinion des consommateurs concernant des attributs sensoriels spécifiques. Ainsi, la majorité des études hédoniques présentent des questions diagnostiques utilisées comme pistes de compréhension des résultats obtenus sur la préférence ou l'appréciation globale. Différents types de questions diagnostiques sont donc souvent ajoutées (figure 5) : des questions ouvertes concernant les raisons d'appréciation ou de non-appréciation, des questions concernant l'appréciation d'attributs spécifiques (appréciations partielles ou notation de l'intensité par rapport à l'idéal) mais également des questions d'opinion (échelles d'accord).

Questions ouvertes
 Qu'est-ce qui vous plaît dans ce produit ? Dites-moi tout ce que vous aimez.
 Qu'est-ce qui vous déplaît ou vous plaît moins dans ce produit ? Dites-moi ce que vous n'aimez pas.

Appréciations partielles
 Pour résumer votre impression sur l'aspect de ce produit, quelle note de 1 à 10 lui donnez-vous ?
 1 signifie que vous en avez une très mauvaise opinion. Au contraire, 10 signifie que vous en avez une très bonne opinion. Les autres notes servent à nuancer votre jugement.

Notation par rapport à l'idéal (JAR)

	BEAUCOUP TROP ...	UN PEU TROP ...	JUSTE BIEN ...	PAS TOUT A FAIT ASSEZ...	PAS DU TOUT ASSEZ...
Ce produit est ... ACIDE					
Ce produit est ... EPAIS					
Ce produit est ... ONCTUEUX					
Ce produit est ... SUCRE					
Le goût de ce produit est ... PRONONCE					

Opinions

	OUI, TOUT A FAIT D'ACCORD	OUI, PLUTOT D'ACCORD	MOYENNEMENT D'ACCORD	NON, PLUTOT PAS D'ACCORD	NON, PAS DU TOUT D'ACCORD
Ce produit a une odeur agréable					
Ce produit est fade					
Ce produit a un arôme naturel					
Ce produit a bon goût					
Ce produit est rafraîchissant					

Figure 5 : Exemples de questions diagnostiques

II.2 - QUEL GROUPE DE SUJETS ?

Le groupe de sujets est l'un des facteurs les plus importants à considérer dans la mise en place d'une étude hédonique. Les études consommateurs ont pour objectif d'aider les développeurs à satisfaire les besoins d'une population cible en particulier. Pour cela, un échantillon de consommateurs représentatif de la population susceptible d'acheter et d'utiliser le produit est sélectionné pour donner son opinion. Cette opinion est ensuite supposée représenter le potentiel de séduction du produit sur le marché. Par ailleurs, l'autre composante essentielle du groupe de sujets dont nous allons discuter concerne le nombre de sujets interviewés qui agit ainsi sur la précision et la stabilité du résultat obtenu.

II.2.1 – Qui interroger ?

Les services marketing fournissent un panorama assez complet des utilisateurs du produit testé en termes de sexe et d'âge. Par ailleurs, même si toutes les personnes interrogées sont généralement des consommateurs de la catégorie de produits testée au sens large, il peut être également important de distinguer, dans le recrutement, les consommateurs potentiels,

occasionnels et réguliers du produit. Cette information est difficile à obtenir a priori puisqu'un long questionnement sur les habitudes d'utilisation d'une marque peut bien évidemment attirer l'attention du sujet sur la marque du produit qui va être dégusté (Cabanis, 2003). Il s'agit donc lors du questionnaire de recrutement d'ajouter des questions sur différentes marques afin de camoufler la marque d'intérêt. Enfin, les études sont souvent réalisées dans des lieux géographiques multiples de manière à ne pas s'intéresser uniquement à la population d'un seul lieu.

D'autres règles sont également importantes à appliquer dans la formation du groupe de sujets. McDaniel et Sawyer (1981) et Shepherd et Griffiths (1987) ont observé un impact du niveau d'expertise des sujets interviewés concernant le produit sur les informations hédoniques récoltées. Ainsi, il n'est pas recommandé d'interroger les employés de l'entreprise qui fabrique le produit. Le nombre de tests auxquels les personnes ont participé est également un critère à prendre en compte. En aucun cas les sujets ne doivent être des « experts » en tests consommateurs. La norme XP V09-500 (AFNOR, 2000) préconise un maximum de vingt participations en un an avec six séances, au plus, portant sur la même famille de produits. Ces contraintes paraissent cependant peu restrictives.

II.2.2 - Combien de sujets ?

La détermination de l'effectif de sujets à interroger présente plusieurs implications sur la mise en place des tests. La première implication est financière étant donné que plus le nombre de sujets à recruter est important, plus le test sera coûteux. La seconde implication concerne la représentativité de l'échantillon qui sera meilleure dans le cas d'un effectif élevé. Enfin, d'un point de vue statistique, l'effectif de l'échantillon influence le pouvoir du test. Ainsi, l'augmentation de l'effectif augmente le pouvoir statistique du test (c'est-à-dire sa capacité à détecter une différence si elle existe). Ce dernier point est à l'origine de plusieurs propositions de tables (AFNOR, 2000 ; Moskowitz *et al.*, 2003) permettant de calculer l'effectif de sujets nécessaire sur la base de la différence critique que l'on souhaite détecter, la variabilité estimée des données et le risque choisi. Cependant, l'utilisation de ces tables peut être risquée puisqu'elles supposent l'adaptation de la taille du panel en fonction de l'objectif de l'étude (souhait de démontrer que les produits dégustés sont appréciés différemment ou au contraire appréciés de la même façon). Cependant, même s'il est plus fréquent d'obtenir des conclusions de non significativité basées sur des échantillons souvent trop faibles, il peut également arriver qu'un échantillon trop important de sujets aboutisse à une différence significative d'appréciation de produits qui ne serait finalement perçue que par une infime

partie des consommateurs (Lawless et Heymann, 1999). Dans l'objectif d'assurer une bonne précision des résultats, la détermination d'un effectif minimum doit être cependant réalisée. La détermination de cet effectif minimum est plutôt controversée. Alors que certains auteurs préconisent l'utilisation de 100 réponses par produit évalué et par segment de population étudié (Cliff *et al.*, 1997 ; Gordon et Norback, 1985 ; Perrien *et al.*, 1984 ; Sharp *et al.*, 1986), d'autres suggèrent un minimum de 60 réponses (AFNOR, 2000 ; Moskowitz, 1997). Un échantillon inférieur à cette taille ne permettant pas de représenter la diversité existant dans les préférences des consommateurs ne conduirait pas à une conclusion stable et exploitable.

Le choix de la taille de l'échantillon reste principalement un compromis entre des risques réels pris par les expérimentateurs et le coût de l'étude. En ce qui concerne le choix de la taille de l'effectif, les utilisateurs de tests consommateurs chez Danone se sont arrêtés sur un compromis scientifique et économique de 120 sujets par cible de consommateurs avec une tolérance jusqu'à 80 sujets.

II.3 - QUELLES CONDITIONS DE TEST ?

Outre le type de réponse hédonique, plusieurs alternatives méthodologiques existent pour collecter l'appréciation des produits. Parmi ces alternatives méthodologiques, nous retrouvons le choix du lieu du test : un laboratoire d'évaluation sensorielle, une salle dans un lieu publique ou le domicile du sujet. La procédure du recueil des réponses peut être également de différente nature et demander une évaluation à l'aide d'un questionnaire auto-administré ou en face-à-face par un enquêteur. Le mode de présentation des échantillons constitue également un point important du protocole des tests consommateurs. Ainsi, nous trouvons différents types de présentations pour lesquelles nous proposons une nomenclature adaptée à leur différenciation qui est souvent peu explicite dans les ouvrages méthodologiques :

- le monadique pur : chaque sujet n'évalue qu'un seul produit (il y a donc autant de groupes de sujets que de produits à évaluer),
- le monadique séquentiel multi-séance : tous les sujets dégustent tous les échantillons présentés lors de séances différentes (un produit par séance),
- le monadique séquentiel mono-séance : tous les sujets dégustent tous les échantillons les uns à la suite des autres dans une même séance,
- le comparatif ou simultané : tous les sujets dégustent l'ensemble des échantillons en simultané dans une même séance.

Bien d'autres caractéristiques du protocole nécessitent généralement un choix de la part de l'expérimentateur : l'ordre de présentation des produits, la quantité de produit présentée, le nombre de produits présentés, le moment de la séance, etc. Ces quelques exemples de choix auxquels un expérimentateur est soumis lors de la mise en place d'un protocole de test consommateurs laissent supposer une grande flexibilité des protocoles. Cependant, en pratique, finalement ce sont principalement les lieux d'évaluation qui conditionnent l'état des autres caractéristiques du protocole. Ainsi, aujourd'hui les instituts d'étude et les universitaires distinguent deux grands types de tests : les tests en salle (ou laboratoire) et les tests à domicile. Nous allons détailler ces deux types de tests et voir qu'ils se différencient sur de nombreux aspects méthodologiques autres que le lieu de l'évaluation.

II.3.1 - Les tests en salle et en laboratoire : les tests en conditions contrôlées

Les méthodologies mises en place lors des tests en salle sont relativement proches de celles mises en place en laboratoire. Globalement, même si ces méthodologies se différencient par le lieu du test, elles présentent toutes deux des conditions de test contrôlées contrairement aux tests à domicile. Alors que les tests en laboratoire sont principalement mis en place par les universitaires, les tests en salle correspondent à la mise en pratique de ces tests en conditions contrôlées par les instituts d'études qui ne possèdent pas de laboratoire. Les salles dans lesquelles sont effectués ces tests en conditions contrôlées sont en général pourvues de plusieurs tables indépendantes les unes des autres (ou de cabines individuelles dans le cas du laboratoire) de façon à éviter la déconcentration du sujet par les autres personnes interviewées en même temps. Les produits anonymes sont préparés à l'avance dans une pièce séparée de la salle d'évaluation et sont présentés les uns après les autres dans des récipients codés et identiques pour tous les produits. La préparation des produits est identique pour tous les échantillons et pour tous les sujets. La séance de dégustation comportant généralement plusieurs échantillons à consommer, les doses présentées sont souvent inférieures aux portions habituellement consommées (sauf pour la procédure en monadique pur). Il est cependant fréquent que les expérimentateurs imposent une quantité minimale de produit à consommer (« buvez au moins la moitié du verre avant de donner votre opinion »). Une fois la dégustation terminée, l'enquêteur demande au sujet de donner son opinion concernant l'appréciation globale du produit puis il lui pose un certain nombre de questions liées à l'appréciation de quelques caractéristiques sensorielles spécifiques ou à son opinion sur des items concernant le produit. Le sujet doit ensuite se rincer la bouche avant de déguster le(s) prochain(s) produit(s) proposé(s) pour le(s)quel(s) il répond aux mêmes questions. Le nombre

de produits présentés lors de ces tests peut être assez élevé mais il est cependant nécessaire de limiter le nombre d'échantillons dégustés lors d'une même séance en fonction de la fatigue physiologique (saturation, satiété) et psychologique des sujets. Ce mode de présentation en monadique séquentiel mono-séance nécessite un ordre de présentation des échantillons équilibré sur l'ensemble du panel interviewé. Il arrive cependant que certaines études soient menées avec une présentation en monadique pur impliquant donc la dégustation d'un seul produit par sujet. Plus rare dans le cas des tests de notation mais possible lors des tests de préférence qui s'intéressent à des différences d'aspect, le sujet peut également comparer tous les produits en même temps avant de donner son évaluation (présentation simultanée des produits).

Dans la littérature, les tests en laboratoire ou en salle se différencient principalement sur deux points : le niveau de contrôle des conditions de dégustation et le panel de sujets interrogés.

II.3.1.1 - Le test en laboratoire

Le test en laboratoire est décrit par Stone et Sidel (1993) comme le test le plus utilisé en raison de sa facilité de mise en œuvre. Ce test est décrit comme permettant la collecte de données issues principalement de résidents locaux car le laboratoire est souvent localisé au sein de l'entreprise ou de l'université. Des instituts peuvent également proposer des études en laboratoire mais les lieux de ces laboratoires sont souvent difficilement accessibles à une population de consommateurs recrutés ponctuellement ce qui nécessite donc un pré-recrutement de sujets panélisés. Les instituts d'études sont aujourd'hui conscients de l'existence de consommateurs qui, attirés par les rémunérations, sont panélisés sur de nombreuses bases de données différentes et qui participent à de nombreuses études. Ainsi, ce type de recrutement sur base de données peut engendrer l'interview de panels peu représentatifs d'une population de consommateurs plus « naïfs ». Ce type de test a cependant l'avantage de présenter le meilleur contrôle de tous les aspects du test, de la préparation des produits aux conditions d'évaluation (la luminosité, la température, le bruit, les distractions dues aux autres interviewés). Stone et Sidel (1993) ajoutent que le strict contrôle des conditions de dégustation ainsi que l'interview de sujets familiers avec ce genre d'expérience permet de n'interroger qu'un faible nombre de sujets pré-recrutés sur une base de données (faible variabilité). En ce qui concerne le nombre de produits à évaluer, l'avantage du pré-recrutement fait que la séance est normalement peu limitée en temps. Bien que Kamen *et al.* (1969) observent que des sujets peuvent évaluer au moins douze produits différents si un

intervalle est bien respecté entre chaque échantillon et que seule l'acceptabilité est mesurée, Resurreccion (1998) ainsi que Stone et Sidel (1993) préconisent un maximum de cinq ou six produits. Par ailleurs, l'avantage du pré-recrutement en laboratoire permet de faire déguster un grand nombre de produits par les mêmes sujets lors de séances différentes. Même si ces tests sont fréquemment mis en place au sein de laboratoires universitaires de recherche, notre expérience des études hédoniques commanditées par les industriels en a que très rarement montré l'utilisation.

II.3.1.2 - Le test en salle

Les salles de test possèdent la particularité d'être localisées dans des lieux de rassemblement d'acheteurs potentiels (rues commerçantes, centres commerciaux). Les sujets sont généralement interceptés et sélectionnés sur place tout au long de la journée par des enquêteurs. Une fois le consommateur sélectionné comme appartenant à la cible, celui-ci est invité à participer à une séance de dégustation qui n'excède pas vingt minutes. Une durée de séance de test supérieure nécessiterait un pré-recrutement et un dédommagement des sujets pour leur participation au test. Ce mode de recrutement « ponctuel » des consommateurs a l'avantage de permettre le recrutement d'un panel de sujets « non experts » et a priori plus représentatif de la cible potentielle du produit. Par ailleurs le lieu facilite le recrutement d'un grand nombre de sujets. Ces sujets sélectionnés ponctuellement ont cependant moins de temps que des sujets pré-recrutés ce qui limite donc le nombre d'échantillons que l'on peut présenter ainsi que le nombre de questions à poser par échantillon. Les tests en salle présentent un moins bon contrôle des conditions de test que les tests en laboratoire. En effet, les salles présentent rarement de vraies cabines individuelles et le va-et-vient incessant des interviewés dans la salle est source de distraction pour les sujets. Les instituts d'études proposent donc généralement des interviews simples en face-à-face, plutôt qu'à l'aide d'un questionnaire auto-administré, permettant de focaliser le sujet sur son interview et non sur celle de son voisin. Ce type de test est donc celui qui est le plus souvent proposé par les instituts d'études étant donné qu'il permet à la fois des conditions de dégustation à peu près contrôlées et l'interview de consommateurs « naïfs » représentatifs de la population cible. Ce dernier point commence aujourd'hui à être remis en question compte tenu du trop grand nombre de salles de tests présentes dans les mêmes endroits. Les habitués de ces lieux sont donc sollicités en permanence et peuvent donc devenir des « experts » en tests hédoniques ce qui diminuerait la validité prédictive des données hédoniques collectées. D'autres tests en conditions contrôlées ont donc été développés et sont décrits dans la littérature bien que leur application en routine

ne soit pas encore très présente. Il s'agit des tests en laboratoire mobile (caravanes) par exemple qui ont l'avantage d'apporter les conditions contrôlées des tests en laboratoire tout en permettant l'interview d'un grand nombre de sujets différents et plus représentatifs.

II.3.2 - Les tests à domicile : les tests en conditions naturelles

Ce type de test a pour objectif de faire évaluer les produits dans des conditions normales d'utilisation. Les sujets sont généralement pré-recrutés sur des bases de données pour représenter la population cible d'intérêt et sont donc souvent dédommagés. En général, les produits, masqués et codés, sont testés un par un pour empêcher le sujet de faire une comparaison directe entre les produits. Par ailleurs, distribuer tous les produits en même temps et demander au sujet de suivre un ordre de dégustation bien défini ajouteraient une source d'erreur incontrôlable. Ainsi, les différents produits sont distribués un par un avec un délai plus ou moins long entre chaque dépôt (de quelques jours à plusieurs semaines). Cette présentation des produits en monadique séquentiel multi-séance est coûteuse en temps ce qui fait que peu de produits différents sont généralement testés lors d'un test à domicile (deux ou trois maximum). Certaines études à domicile présentent également les produits en monadique pur où chaque sujet ne déguste qu'un seul produit. Le mode de distribution des produits est généralement dicté par le budget du test. Ainsi, trois solutions sont envisageables : un dépôt au domicile des sujets (le plus coûteux et le moins risqué), un envoi par la poste (par toujours possible et risqué) ou un déplacement des sujets (le moins coûteux s'il existe un lieu de stockage). Le premier produit est distribué en quantité importante et accompagné d'une notice contenant des instructions de dégustation plus ou moins précises. Le consommateur doit utiliser le produit pendant une durée déterminée comme il le souhaite et souvent à la fréquence qu'il le désire. Son opinion sur le produit peut être recueillie de différentes façons. Un enquêteur lors d'une seconde visite peut interviewer le sujet sur le produit qu'il vient de tester pendant ces quelques jours (appréciation globale puis questions diagnostiques). L'enquêteur peut également récupérer un questionnaire auto-administré distribué préalablement dans lequel le sujet a reporté son évaluation au bout de plusieurs consommations du produit. Dans le cas de l'évaluation de plusieurs produits, lors de cette seconde visite l'enquêteur distribue les échantillons du deuxième produit testé avec les mêmes instructions.

Daillant-Spinnler et Issanchou (1995) et Murphy *et al.* (1958) décrivent l'utilisation d'une méthode de test appelée « In-home test » qui consiste à mettre en place une dégustation des produits en monadique séquentiel mono-séance au domicile de l'interviewé. Cependant

cette méthode de test proposant une seule consommation ponctuelle des produits dans un environnement naturel n'est décrite dans aucun manuel méthodologique.

B - IMPLICATION DES VARIABLES DE CONTEXTE DANS LA FORMATION DU JUGEMENT HEDONIQUE

Nous venons de décrire les enjeux de la mesure hédonique dans un contexte industriel agroalimentaire ainsi que l'état des pratiques concernant sa collecte. La figure 3, présentée précédemment, sur les déterminants du comportement alimentaire illustre l'importance de la composante attitudinale dans le choix alimentaire. Les notes hédoniques ou préférences émises par les sujets lors des tests organoleptiques traduisent leur opinion envers l'aliment testé. L'attitude envers les propriétés sensorielles de l'aliment est donc ce que les tests hédoniques en aveugle tentent de récupérer. Il convient donc d'explorer les fondements de cette notion d'attitude (définition et formation) que nous nous employons à mesurer par l'intermédiaire d'une réponse évaluative à l'égard d'un objet. Par ailleurs, la description de la formation de l'attitude nous permettra d'introduire le rôle du contexte d'évaluation dans le processus d'émission d'une attitude. En effet, les informations relatives à l'objet cible ne sont pas les seules données à partir desquelles le jugement est émis. Dans le cadre des préférences alimentaires, nous observons sur la figure 6 que de nombreuses autres informations non directement liées à l'aliment (la cible) exercent une influence souvent non négligeable dans le déterminisme des préférences, et peuvent être rassemblées sous le terme commun de « contexte ». Nous détaillerons donc le rôle de ce contexte d'évaluation dans la formation d'un jugement hédonique d'un consommateur face à un aliment.

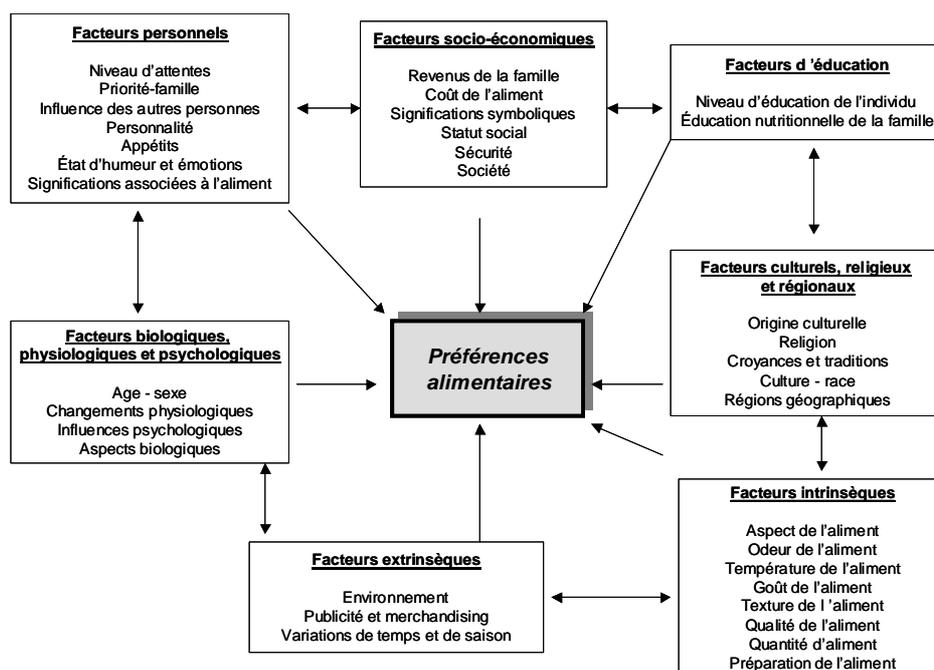


Figure 6 : Facteurs influençant le jugement de préférence (d'après Khan, 1981)

I - LE JUGEMENT EVALUATIF : PROCESSUS INTEGRATIF A LA BASE DE L'ATTITUDE

Le jugement d'un objet, d'une personne est un acte que nous réalisons de manière coutumière sans en avoir conscience et permet d'aboutir à une attitude envers cet objet. Après avoir défini et détaillé le processus à la base de la formation de l'attitude au sens large, nous expliquerons comment celui-ci s'applique à la formation du jugement hédonique d'un aliment.

I.1 - L'ATTITUDE : DEFINITION ET PROCESSUS DE CONSTRUCTION

Dès 1935, Allport définissait l'attitude comme « un état mental de préparation, organisé au travers de l'expérience, exerçant une influence directrice ou dynamique sur la réponse de l'individu à tous les objets et situations avec lesquels il est en relation ». L'attitude est donc unanimement définie comme un élément cognitif traduisant une position plus ou moins favorable de l'individu vis-à-vis de l'objet et l'orientant vers certaines actions particulières. Ainsi, ce rôle de l'attitude dans le comportement humain explique que de nombreuses disciplines (psychologie, sociologie, marketing) s'attachent à la mesurer. La définition conceptuelle de l'attitude semble quant à elle moins consensuelle dans la littérature avec deux approches distinctes.

L'attitude a été envisagée tout d'abord comme une instance organisée et durable au même titre que les concepts de soi, de personnalité ou de caractère. D'ailleurs, en 1935, Allport déclarait que les attitudes « sont souvent aussi rigides que les habitudes » et « persistent souvent tout au long de la vie une fois qu'elles ont été fixées pendant l'enfance ou l'adolescence ». Le modèle traditionnel envisage donc l'attitude comme un fichier mental dont l'individu récupère le contenu à chaque fois qu'il est sollicité pour exprimer son attitude. Cependant ce modèle basé sur une attitude stable et stockée en mémoire rend difficile la compréhension de la variabilité des jugements évaluatifs. Ainsi, une deuxième approche conceptuelle de l'attitude suggère que les attitudes sont plus le fruit d'un processus de construction que d'un processus de récupération. Certains chercheurs supportent donc la thèse que plutôt que de récupérer un fichier mental, les individus construiront épisodiquement leur attitude sur la base d'informations temporairement saillantes ou accessibles. La définition de l'attitude donnée par Eagly et Chaiken (1993) comme étant « une tendance psychologique qui s'exprime par l'évaluation d'une entité particulière avec un certain degré de faveur ou de défaveur » met l'accent sur l'évaluation. Ainsi, un consommateur n'a pas d'attitude tant qu'il

n'évalue pas une entité. D'après Bonin (2003), cette perspective constructiviste implique plusieurs postulats :

- Il n'y a pas d'attitude définitive.
- Il n'y a pas une attitude envers un objet mais plutôt un certain nombre d'attitudes dépendant du nombre d'informations disponibles pour penser à l'objet (Tesser, 1978).
- Le jugement évaluatif est produit au moment de la rencontre avec l'objet. Ce sont les informations à l'origine de l'attitude (sentiments, croyance et comportements passés) qui sont récupérées et non l'attitude elle-même. Dans certains cas, les jugements évaluatifs passés peuvent avoir été stockés en mémoire et donc être utilisés pour construire la nouvelle attitude au même titre que les autres informations.
- Ce processus de construction est largement dépendant à la fois de processus introspectifs internes et du contexte externe dans lequel l'attitude est exprimée. Ainsi, au contraire de l'approche de l'attitude comme entité durable, l'attitude est largement soumise aux effets de contexte.

I.2 - PROCESSUS DE CONSTRUCTION DU JUGEMENT HEDONIQUE D'UN ALIMENT

Les mécanismes sous-jacents à la formation d'un jugement évaluatif s'appliquent également à la formation de la perception hédonique alimentaire. En effet, de la même façon et de manière tout aussi inconsciente lors d'un acte alimentaire, les données sensorielles sont traduites en perception hédonique par le biais d'un jugement évaluatif. Ce jugement hédonique est indispensable à l'être humain afin de lui faciliter sa prise de décision concernant l'acte de re-consommer ou non un aliment auquel il est confronté.

Dans ses travaux pour définir la qualité sensorielle d'un aliment, Williams (1983) relie les caractéristiques sensorielles de l'aliment non seulement aux paramètres physiques mais également à la préférence du consommateur. En faisant cela, il met en évidence la relation entre le stimulus, l'organisme et la réponse. La qualité sensorielle peut alors être définie comme le degré d'appréciation ou de rejet des caractéristiques sensorielles que le consommateur perçoit. Ainsi, même si le plaisir peut être défini comme une sensation sensorielle inconsciente (Chiva, 1992 ; Pieron, 1974), ce message sensoriel est ensuite interprété par le sujet et peut être verbalisé de manière tout à fait consciente sous la forme « j'aime » ou « je n'aime pas ». Chiva (1992) décrit la perception hédonique d'un aliment comme la lecture et le sens donné au message sensoriel. Par ailleurs, durant notre vie, nous acquérons, stockons et actualisons une vaste quantité d'informations à propos des aliments. A

chaque occasion de consommation, une partie de cette information stockée est activée et ajoutée aux nouvelles informations pour déterminer un comportement. Ainsi, la perception hédonique d'un aliment est donc « une construction mentale, issue de l'expérience personnelle, qui permet de donner un sens aux messages sensoriels, de les associer dans une perception plus large et de les rendre signifiants ». McEwan et Thomson (1988) reprennent le modèle proposé par Olson (1981) et structurent les principaux processus cognitifs inconscients qui entrent en jeu dans le comportement alimentaire autour du paradigme Stimulus-Organisme-Réponse (figure 7).

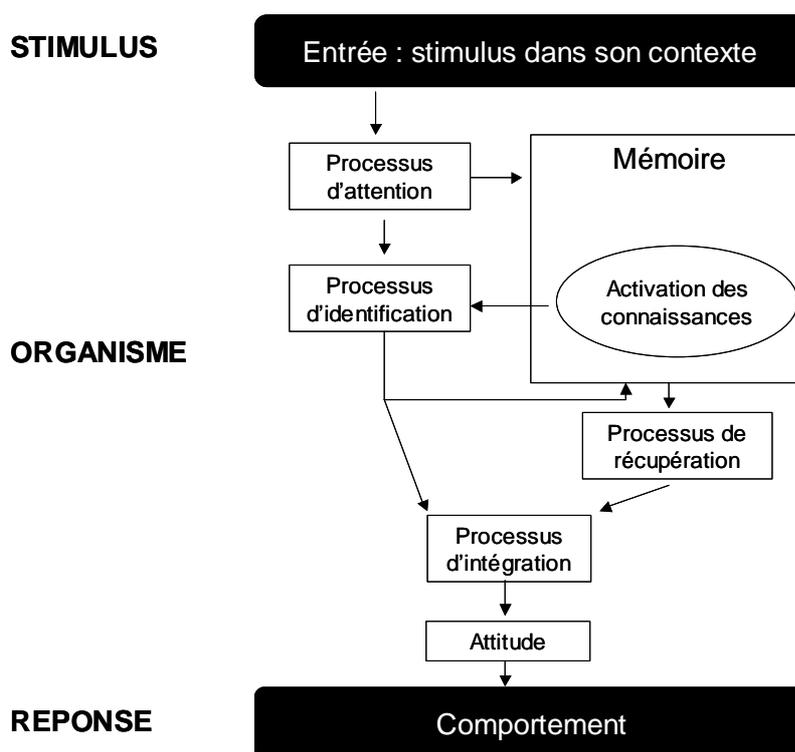


Figure 7 : Processus cognitifs qui entrent en jeu dans le comportement alimentaire (d'après Olson, 1981)

L'organisme est face à des entrées sensorielles (existence du stimulus dans son contexte) dont l'attention nécessite un effort cognitif qui active des connaissances stockées en mémoire concernant le stimulus testé. Il s'en suit une étape d'identification pendant laquelle les informations apportées par le stimulus dans son contexte sont interprétées et identifiées à la lumière des connaissances existantes qui ont été spécifiquement activées en mémoire. L'identification dépendra donc clairement des significations acquises précédemment et le résultat est une interprétation psychologique directe des informations physiques. Si le stimulus est complètement nouveau, son identification peut se faire sans l'intervention de la mémoire mais cette identification sera automatiquement mémorisée et réutilisée lors d'une

consommation future du même aliment. La dernière étape consiste en l'intégration des informations précédemment obtenues sur l'identification du stimulus et sur les connaissances stockées en mémoire et activées (croyances générales, affectif, images visuelles, valeurs, expériences personnelles, etc). Ainsi, le consommateur fait son jugement de valeurs :

- Est-ce qu'il aime les caractéristiques sensorielles ?
- Est-ce qu'il pense que c'est efficace nutritionnellement ?
- Est-ce qu'il pense que c'est cher ?
- Etc.

Ces évaluations détermineront ensuite son attitude envers le stimulus et donc son comportement en réponse au stimulus (le produit dans son contexte).

II - L'INFLUENCE DU CONTEXTE DANS CE PROCESSUS INTEGRATIF

II.1 - DEFINITION

Un grand nombre de travaux ont montré que la quasi-totalité des comportements humains sont influencés par les variables de contexte et le comportement alimentaire ne déroge pas à cette règle. En effet, comme nous venons de le voir, la décision de consommer un aliment est le résultat d'un processus psychologique complexe qui débute par l'expérience du stimulus dans son contexte. Le contexte est un concept très large et pas toujours très bien défini. Le grand dictionnaire de la psychologie présente le contexte comme la situation au sein de laquelle se trouve un stimulus. De la même façon, Rozin et Tuorila (1993) définissent le contexte comme « l'ensemble des évènements et expériences qui ne font pas partie de l'événement de référence lui-même mais qui sont en relation avec lui ». Le contexte dans lequel le comportement est émis a en fait été identifié comme générateur d'informations indispensables à la construction du comportement. Ainsi, compte tenu de la relation entre le contexte et l'événement de référence, le contexte peut être défini comme l'ensemble des informations (conscientes et inconscientes) non directement liées à l'objet cible et exerçant une influence sur l'événement de référence. Selon le domaine d'étude, nous avons pu remarquer que l'entité du contexte pouvait être plus ou moins large et dépend finalement de la cible étudiée. Bonin *et al.* (2001) définissent le contexte comme « tout ce qui n'est pas la cible ». Il s'agit donc de l'ensemble des informations qui accompagnent la réalisation d'un événement cible ou la production d'une réponse cible. Cette définition a l'avantage d'être applicable à n'importe quel domaine d'étude.

Les tests consommateurs réalisés en évaluation sensorielle ont pour objectif de mesurer l'appréciation d'un aliment uniquement focalisé sur les propriétés intrinsèques de ce dernier c'est-à-dire sur la recette. Cependant, compte tenu du nombre d'informations qui entrent en jeu dans la formation de ce jugement évaluatif, il semble difficile de déterminer dans une mesure hédonique ce qui est la part apportée par les seuls stimuli sensoriels de l'aliment. Ainsi, nous avons choisi de définir dans ce travail de thèse comme variables de contexte toutes les variables qui influencent le jugement hédonique récolté autres que les variables intrinsèques à l'aliment testé à savoir ses propriétés sensorielles.

II.2 - LES EFFETS DE CONTEXTE DANS LES ENQUÊTES ATTITUDINALES

La recherche dans le domaine des enquêtes attitudinales (enquêtes visant à récolter l'opinion des individus sur un thème/objet spécifique) s'est intéressée à l'influence du contexte sur les jugements évaluatifs. En effet, il n'est pas rare d'observer que les gens produisent des évaluations différentes envers un même objet à diverses occasions, alors que cet objet n'a subi aucun changement entre les différents moments d'évaluation. Une explication probable de l'inconsistance parfois observée entre les évaluations répétées d'un même objet réside dans le changement des influences contextuelles implicites s'exerçant sur l'individu lors des différentes occasions d'évaluation. Savoir comment ces influences sont possibles, par quels mécanismes elles sont susceptibles d'émerger et quelles en sont les conséquences sont autant de questions auxquelles nous allons tenter de répondre.

II.2.1 - Par quels mécanismes le contexte influence-t-il le jugement ?

Les effets de contexte s'expliqueraient principalement par une modification des processus d'encodage (Bower *et al.*, 1981), de récupération (Tourangeau *et al.*, 1989) et d'intégration (Schuman et Presser, 1981) des informations intervenant dans la formation du jugement. Comme nous l'avons vu précédemment, les réponses à des questions d'attitude sont habituellement basées sur une attitude générée au moment où la question est posée. Bien que certains sujets puissent avoir stocké en mémoire une évaluation préexistante du problème d'attitude sur laquelle leurs réponses sont basées, beaucoup d'autres ne récupèrent pas simplement une réponse toute prête. Les jugements sont donc produits au moment où la question est posée mais sont basés, pour une grande partie, sur un matériel stocké dans la mémoire à long terme. Bien évidemment, une fois qu'un jugement a été généré, il sera aussi stocké en mémoire et pourra être utilisé ultérieurement.

Lors d'enquêtes attitudinales, la génération des réponses aux questions d'attitude n'est pas forcément basée sur un schéma systématique de toutes les informations appropriées. En effet, compte tenu du faible délai dont dispose le sujet, la génération des réponses est plutôt basée sur un rapide échantillonnage de ces informations. Ainsi, la composition de cet échantillonnage peut varier d'une occasion à l'autre et les réponses peuvent donc changer. D'après Feldman et Lynch (1988) et Moorthy (1991) la quantité et le type d'informations récupérées et utilisées pour aboutir au jugement global d'un objet dépend :

- de la disponibilité des informations en mémoire,
- du diagnostic relatif de ces informations disponibles,
- de l'accessibilité relative de ces informations disponibles.

Moorthy (1991) donne quelques exemples de chacune de ces caractéristiques qui entrent en jeu dans l'échantillonnage des informations utilisées pour construire un jugement global.

La disponibilité des informations joue un rôle principalement sur la quantité d'informations récupérées et utilisées. Cela peut par exemple se traduire en termes « d'expertise » de l'interviewé concernant la catégorie de l'objet évalué : plus le sujet interviewé est expert plus la probabilité d'avoir des informations disponibles est forte. Dans le cas de l'évaluation d'un aliment lors des tests organoleptiques, cela se traduit par la différence qui est faite entre un consommateur « naïf » et un sujet plus expert. Nous comprenons donc mieux le souhait de ne pas interroger des consommateurs experts d'une catégorie de produits afin de ne pas récolter un jugement basé sur des informations trop nombreuses par rapport à un jugement basé sur un échantillonnage d'informations effectué par des consommateurs plus « naïfs ».

Le diagnostic d'une information a une influence sur la sélection de cette information parmi d'autres étant donné qu'il conditionne généralement l'importance pour le sujet de l'information dans la construction de son jugement global. Il dépend par exemple de la qualité de cette information. Plus le sujet est certain de cette information, plus il fera appel à elle. Plus l'information semble pertinente, plus elle sera utilisée. Un jugement global peut également faire appel à une opinion comparative d'un objet. Dans ce cas, une information importante pourra ne pas être utilisée si le sujet pense que cette information n'est pas discriminante. Ainsi, un jugement global peut se construire en mettant de côté des informations positives sur les produits si ceux-ci se différencient uniquement sur des informations négatives. La façon dont le sujet voit la tâche du jugement peut également influencer le diagnostic de l'information et donc son utilisation. Par exemple, si le sujet veut faire plaisir à l'enquêteur, il peut ne sélectionner que des informations positives, ce qui peut aboutir à un jugement globalement plus favorable que ce qu'il perçoit réellement (désirabilité sociale).

Enfin, **l'accessibilité** d'une information, au même titre que le diagnostic, détermine si celle-ci sera utilisée dans la construction du jugement global. L'accessibilité d'une information dépend à la fois de facteurs à long terme (son importance et sa fréquence d'utilisation) et de facteurs à court terme (son activation récente, sa pertinence ou sa relation avec d'autres informations déjà activées).

Par ailleurs, le nombre d'informations considérées (disponibilité) ainsi que leur sélection est fonction de la motivation et des conditions de l'intégration de ces informations. Plus les conditions d'intégration seront perturbées (délai bref, distractions) et moins le niveau

de diagnostic entrera en jeu dans la sélection des informations. Hoch (1984) précise en effet que la caractéristique d'accessibilité de l'information pré-value sur la caractéristique de diagnostic de l'information. Ainsi, une information récemment activée mais non pertinente risque d'inhiber le rappel et l'utilisation d'une information pertinente mais moins accessible dans le jugement global. D'après Tourangeau *et al.* (1989), c'est donc principalement la composante de récupération des informations qui va être influencée (et donc biaisée) par les variables contextuelles lors de la génération d'un jugement global. Cette tendance à baser son jugement sur les informations les plus accessibles a été appelée disponibilité heuristique par Tversky et Kahneman (1973).

II.2.2 - Comment les effets de contexte se manifestent-ils ?

Dans le domaine des enquêtes attitudinales, deux types d'effets de contexte ont été mis en évidence : l'assimilation et le contraste. Ces deux types d'effets résultent du degré de chevauchement entre les caractéristiques des informations contextuelles et les caractéristiques du stimulus à évaluer.

Un effet d'assimilation est observé lorsque qu'il existe une relation forte positive entre les jugements de valeur des stimuli contextuels et le jugement de valeur alloué à la cible. Ainsi, dans le cadre de l'appréciation d'un aliment, il y aura assimilation si la présence d'une personne ou d'une information jugée positivement amène le consommateur à juger l'aliment plus favorablement que si ces informations contextuelles n'avaient pas été activées. De la même façon, le jugement donné par un consommateur dans un état d'humeur négatif sera susceptible d'être soumis à un effet d'assimilation si son jugement concernant l'aliment testé en est défavorisé. L'effet d'assimilation est donc le résultat d'une diffusion du jugement de valeurs du contexte vers le jugement de la cible dans le but de réduire « l'inconfort mental » du sujet (Deliza et MacFie, 1996).

Les effets de contraste sont le résultat d'un scénario inverse c'est-à-dire lorsque les valeurs du contexte et de la cible ne sont pas chevauchantes. Un effet de contraste est observé lorsque le jugement ne reflète aucune relation ou une relation strictement négative entre la valeur allouée à l'objet cible et les valeurs allouées aux stimuli contextuels qui accompagnent la cible. Ainsi, une information contextuelle positive peut entraîner un jugement négatif de la cible et inversement. Le rôle des attentes dans ces effets de contraste est bien connu. Imaginons que le contexte de consommation d'un aliment soit particulièrement favorable (un menu onéreux dans un grand restaurant par exemple). L'attente que construit le consommateur avant de déguster le plat est alors probablement très forte. Pendant la

dégustation de l'aliment, cette attente positive va entrer en jeu comme variable contextuelle dans la construction du jugement du consommateur. Ainsi, si l'aliment est finalement jugé moins positivement que ce que le consommateur attendait, ce dernier aura tendance à sous-évaluer l'aliment dégusté dans ce restaurant par rapport à l'évaluation qu'il en aurait faite dans un environnement moins favorable.

II.3 - TYPOLOGIE DES VARIABLES DE CONTEXTE DANS LE CADRE DE LA PERCEPTION HEDONIQUE D'UN ALIMENT

Au vu du mécanisme intégratif à la base de la formation de la perception hédonique d'un aliment, il convient à ce moment de notre exposé d'identifier la nature des informations auxquelles le sujet est soumis de manière souvent inconsciente lorsqu'il construit sa perception hédonique. D'après les définitions que nous avons pu trouver, le contexte est de manière consensuelle considéré comme un ensemble. Nous allons nous attacher maintenant à décrire la façon dont cet ensemble est structuré dans la littérature.

II.3.1 - Typologie des variables de contexte selon leur source

Si nous considérons les variables de contexte comme étant porteuses des informations qui seront par la suite intégrées dans le processus de génération du jugement, il semble naturel que les toutes premières propositions de classification des variables de contexte (Meiselman *et al.*, 1988) dans le domaine du comportement alimentaire se basent sur les trois sources possibles d'information : le produit, l'individu et l'environnement. Ainsi, nous retrouvons, dans la classification des variables de contexte pouvant influencer la perception hédonique, les trois composants identifiés comme déterminants du comportement alimentaire (Pilgrim, 1957 ; Shepherd, 1985 ; Steenkamp, 1993). En effet, au début du 20^{ème} siècle, la majorité des recherches et avancées dans les domaines de la technologie des procédés agroalimentaires et le développement de produits sont basées principalement sur les caractéristiques du produit en lui-même. Le produit doit convenir à des caractéristiques (sensorielles, nutritionnelles, coûts de production...) établies par l'entreprise. Dans les années 60 commence à émerger le rôle du consommateur dans l'acceptabilité du produit (consumérisme). Ainsi, des travaux de recherches basés sur la psychologie, la physiologie et la sociologie sont réalisés pour tenter de mieux comprendre les déterminants du choix alimentaire. Enfin, dans les années 80 les chercheurs travaillant sur le comportement alimentaire proposent de prendre en compte dans la mise en place de leurs études les variables environnementales ou situationnelles (physiques, sociales, économiques...).

Bonin *et al.* (2001) déclarent qu'un jugement évaluatif formulé à l'égard d'un aliment résulte du traitement de différents types d'informations classées également selon leur source :

- des informations concernant les propriétés organoleptiques du produit,
- des informations concernant les connaissances et représentations construites et enregistrées par le sujet,
- des informations contextuelles associées aux occasions de consommation de l'aliment.

De manière plus simple, Cardello (1995) propose une catégorisation des variables de contexte en deux groupes :

- les variables liées à l'aliment qui incluent les autres aliments servis avec l'aliment cible, les aliments et boissons consommés récemment, etc,
- les variables non reliées à l'aliment qui représentent plutôt l'environnement social, les conditions physiques, le moment de la journée, etc.

II.3.2 - Typologie des variables de contexte selon leur caractéristique « interne » ou « externe »

Bonin (2003) parle d'informations contextuelles externes et internes. Les informations contextuelles externes correspondent à toutes les informations extérieures à la cible et à l'individu présentes dans l'environnement quand l'individu traite la cible. Les informations contextuelles internes quant à elles renvoient à l'ensemble des construits psychologiques (connaissances et représentations) mobilisés par l'individu dans la situation. Enfin, elle ajoute une autre catégorie de variables de contexte qu'elle définit comme les informations contextuelles intrinsèques. Ces dernières réfèrent aux aspects secondaires de la cible qui sont nécessairement perçues avec elle. Par exemple, dans le cas des tests hédoniques en aveugle, la forme et la couleur du pot d'un yaourt (même masqué) sont des informations contextuelles intrinsèques par rapport à la cible (les caractéristiques organoleptiques du produit). Ces variables de contexte intrinsèques s'opposent donc aux variables de contextes extrinsèques décrites précédemment qui présentent la particularité, au contraire, d'entourer la cible mais de ne pas être indispensables à son traitement. Ainsi, s'affranchir de l'influence de ces variables de contexte intrinsèques est impossible. Drifford *et al.* (1995) complètent cette définition en associant aux variables de contexte externes un caractère visible et aux variables de contexte internes un caractère invisible. Dans un vocabulaire plus adapté à la mesure de la perception hédonique d'un aliment, Issanchou et Hossenlopp (1992) parlent quant à eux de variables « contrôlées » ou « non contrôlées ».

II.3.3 - Typologie des variables de contexte selon leur caractéristique temporelle

La lecture des travaux de Rozin et Tuorila (1993) montre un choix de classer les variables de contexte en facteurs « simultanés » et « temporels ». Les facteurs simultanés sont définis comme les variables contextuelles physiquement présentes durant l'événement de référence et les facteurs temporels comme les variables contextuelles provenant d'événements passés ou futurs qui sont activés par le sujet au moment ou l'événement de référence arrive.

De la même façon, Issanchou et Hossenlopp (1992) ont intégré la notion de temporalité dans leur manière d'organiser les informations qui entrent en jeu lors d'un acte de consommation pour aboutir à la construction de la perception hédonique. Ainsi, le sujet évalue son plaisir à un moment donné en faisant appel à des informations liées :

- au passé, au vécu alimentaire (expériences acquises et mémorisées de l'aliment présenté à la dégustation),
- au présent (perception du stimulus dans le contexte du moment),
- au futur (anticipation des effets attendus et prévention sur les événements à venir).

II.3.4 - Vers une globalisation des variables de contexte

Drifford *et al.* (1995) ajoutent à la catégorisation en variables externes et internes une forte composante d'interaction entre ces deux types de variables de contexte en définissant la notion de contexte comme une entité englobant à la fois :

- les connaissances et les représentations construites et enregistrées en mémoire par l'individu sur les produits alimentaires,
- la situation dans laquelle l'individu est amené à consommer le produit,
- la relation existant entre les cognitions et la situation dans la mesure où la situation fournit à l'individu des indices spécifiques de remémoration.

De la même façon, Bernard *et al.* (1993) proposent de considérer dans le processus intégratif de formation d'un jugement la relation entre les situations passées qui interviennent dans l'élaboration des cognitions des individus et les situations actuelles qui interviennent dans la mobilisation de ces cognitions. En effet, les travaux de Tulving (1976) attestent de l'existence d'une mémoire épisodique qui réfère au stockage des expériences passées en termes de dispositions et d'attentes dans la mémoire à long terme. Ainsi, lorsqu'un individu juge un produit, il le juge en fonction de sa propre expérience alimentaire. Il est en effet généralement admis qu'à la simple vue d'un produit nous sommes capables d'activer très rapidement le nom de ce produit, son goût et ses situations habituelles de consommation. Tous ces événements, ces épisodes sont traités par l'individu, encodés et transformés en une

trace en mémoire. Cette trace est d'autant plus accessible que les indices ayant servi à son encodage sont les mêmes que ceux présents dans la situation de remémoration. Le consommateur doit donc être considéré comme un être ayant en mémoire un « historique alimentaire » et possédant des traces qui correspondent à des épisodes de consommation spécifiques à chaque produit et réactivés dans des situations bien spécifiques.

Dans un même souci de globalisation des variables de contexte, Bell et Meiselman (1995) déclarent qu'« un choix dans une situation de consommation est déterminé par la combinaison de l'état de l'environnement avant et pendant l'expérience de choix et l'état de l'individu avant et pendant l'expérience de choix ». Ainsi il semble clair que malgré un souhait de catégoriser de manière très distincte les variables de contexte, leur interaction est constante. Mathieu (2002) précise que la définition de la notion de contexte se heurte à la difficulté de l'identifier, dans la mesure où le contexte est d'abord le résultat d'une interaction entre éléments de différentes natures : les systèmes mémoriels, l'état de l'individu (son histoire, sa personnalité, ses expériences, etc.) et la situation dans laquelle il expérimente un objet. Le schéma proposé par Bonin *et al.* (2001) récapitule parfaitement les différentes propositions de catégorisation des variables de contextes qui influencent le comportement d'un individu face à un aliment (figure 8).

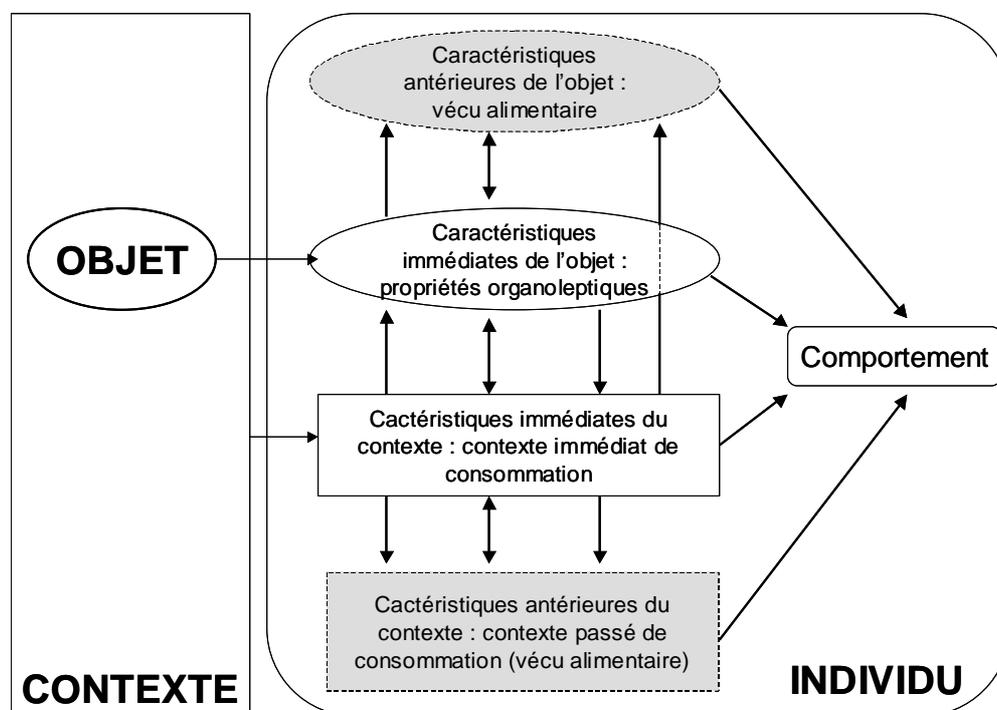


Figure 8 : Schéma récapitulatif des différentes catégories de variables de contexte qui entrent en jeu dans le comportement d'un consommateur face à un aliment (Bonin *et al.*, 2001)

Ce schéma illustre bien que les différentes typologies précédemment décrites ne sont pas exclusives les unes des autres et peuvent être combinées. Ainsi, nous retrouvons la typologie selon la source des informations (l'objet, l'individu et le contexte) combinée à la typologie selon les caractéristiques temporelles (variables simultanées en blanc et variables temporelles en gris). Enfin, ce schéma illustre également l'existence d'interactions entre les différentes variables de contexte.

C - MISE EN EVIDENCE DE L'INFLUENCE DU CONTEXTE SUR LA MESURE HEDONIQUE

Cette partie a pour objectif de faire le point sur les travaux réalisés dans le but d'étudier l'influence du contexte de consommation (au sens large) sur la mesure hédonique. La nature des données hédoniques recueillies dans les travaux présentés dans ce chapitre est variable. Il peut s'agir de données déclaratives comme une note ou une préférence ou de données comportementales comme un choix ou une quantité ingérée.

Nous constatons qu'il existe deux grands types d'approches que nous allons explorer de manière distincte :

- les travaux qui comparent des mesures hédoniques obtenues dans des environnements de consommation globalement différents,
- les travaux qui étudient plus précisément l'impact d'une ou plusieurs variables contrôlées sur la mesure hédonique.

I - INFLUENCE DE L'ENVIRONNEMENT GLOBAL DE CONSOMMATION SUR LA MESURE HEDONIQUE

Il nous a semblé important de distinguer et d'aborder séparément :

- les études qui comparent des mesures hédoniques obtenues dans différents environnements naturels de consommation (lieux publics, domicile),
- et les études qui comparent des mesures hédoniques obtenues par l'intermédiaire de situations de test soit dans des environnements naturels de consommation (lieu public, domicile) soit dans des environnements non naturels de consommation (laboratoire, salle de test).

I.1 – UNE MESURE HEDONIQUE SENSIBLE AU LIEU DE CONSOMMATION

I.1.1 - Impact systématique du lieu de consommation sur la mesure hédonique

La modification de l'appréciation d'un même aliment consommé dans des lieux différents a déjà été observée depuis plusieurs décennies (Green et Butts, 1945 ; Maller *et al.*, 1980 ; Miller *et al.*, 1955). Cependant ce sont les résultats obtenus par l'équipe de Natick (Meiselman *et al.*, 1988) concernant la différence des comportements alimentaires de militaires lors de manœuvres ou dans une cafétéria qui ont réellement initié l'engouement

pour la prise en compte des variables de contexte dans l'étude du comportement alimentaire. Cette même équipe de chercheurs a, par la suite, mis en place d'autres travaux mettant en jeu des individus et des environnements de consommation plus courants. Ils ont alors également mis en évidence une influence de l'environnement de consommation sur l'appréciation d'un même plat ou d'un même repas servi dans des lieux publics différents (Edwards *et al.*, 2003 ; Meiselman *et al.*, 2000). L'effet du lieu de test sur l'appréciation globale est largement significatif et les auteurs retrouvent une hiérarchisation des notes d'appréciation en fonction du prestige des lieux. Ainsi, les notes les plus élevées sont données dans les restaurants prestigieux et les notes les plus basses dans des lieux comme un camp militaire ou un restaurant d'entreprise.

1.1.2 - Comment le lieu de consommation influence-t-il la mesure hédonique ?

L'interprétation des résultats concernant la comparaison des comportements alimentaires de militaires en manœuvre ou dans une cafétéria (Meiselman *et al.*, 1988) fait clairement intervenir le rôle des variables de contexte. Ainsi, la monotonie du régime alimentaire en manœuvre ou encore la qualité de l'état physique et psychique des militaires expliqueraient que ceux-ci consomment moins de ration lorsqu'ils sont en manœuvre que lorsqu'ils sont dans une cafétéria. De la même façon la situation de consommation composée d'un environnement social restreint, d'un confort limité (pas de table ni de chaises) ou encore de l'impossibilité de pouvoir réchauffer la ration pourrait expliquer cette différence de comportement.

Dans leur étude, Edwards *et al.* (2003) interprètent la variabilité des performances du produit consommé en fonction du lieu par les attentes véhiculées par chacun des lieux. En effet, les consommateurs ont des attentes très claires en ce qui concerne la qualité des aliments proposés dans les différents lieux de consommation (Cardello *et al.*, 1996). L'attente est définie par Cardello (1994a) comme la croyance que les produits seront appréciés ou non appréciés à un certain degré. L'attente hédonique est donc une attitude, une opinion qui se construit en général sur la base des informations contextuelles. Elle entre en jeu au même titre que les autres jugements de valeur dans la construction du jugement final du produit. Lorsque la performance du produit une fois dégusté est équivalente aux attentes celles-ci n'ont, a priori, pas d'influence sur l'appréciation globale. En revanche quand la performance du produit n'est pas équivalente aux attentes, le rôle joué par les attentes (contextuelles) dans le jugement global peut de nouveau s'expliquer par les deux types d'effets de contexte présentés précédemment. Un effet d'assimilation peut se mettre en place et les notes d'appréciation

bougent alors dans le sens des attentes contextuelles, qu'elles soient positives ou négatives. Un effet de contraste peut également être observé et une personne déçue par un produit pourra très sévèrement le sous noter par rapport à l'appréciation qu'il en a. Cependant, d'après Cardello et Sawyer (1992), les effets de contrastes sont rarement observés [voir Cardello (1994a) pour une revue plus détaillée des différents modèles explicatifs de l'influence des attentes]. Les résultats de l'étude d'Edwards *et al.* (2003) illustrent quant à eux plutôt un effet d'assimilation étant donné que le produit est jugé dans le sens des attentes véhiculées par les lieux.

Contrairement aux études que nous allons décrire par la suite, l'avantage de ces études réalisées dans des lieux publics est qu'en général les personnes ne sont pas recrutées spécifiquement pour participer aux études. Elles sont interviewées a posteriori uniquement si elles ont choisi de consommer le produit d'intérêt. Ainsi, nous pouvons imaginer que leur état interne (physiologique, physique et psychologique) est très proche de celui d'une situation naturelle de consommation ce qui nous laisse supposer que leurs réponses ne sont pas biaisées par la situation de test. En revanche, cette méthodologie possède l'énorme inconvénient de n'interviewer que les populations habituelles de ces lieux. L'influence du lieu de consommation est donc largement confondue avec l'influence des caractéristiques socio-démographiques des sujets. L'étude d'Edwards *et al.* (2003) a en effet mis en évidence une influence du lieu de consommation mais également de l'âge des sujets sur le niveau d'appréciation de l'aliment. Les jeunes semblent noter plus sévèrement le produit que les personnes plus âgées ce qui finalement pourrait expliquer pourquoi les notes dans les lieux peu prestigieux sont moins élevées que les notes dans les restaurants luxueux.

L'ensemble de ces travaux sur l'influence du lieu de test illustre la complexité du problème lié à la généralisation d'une mesure hédonique obtenue dans un certain environnement à un autre environnement.

Tableau 1 : Résultats des études ayant comparé des mesures hédoniques obtenues en conditions artificielles ou naturelles de consommation

Etude	Produits testés	Protocole en situation artificielle	Protocole en situation naturelle	Principales conclusions concernant la comparaison des méthodes
Miller <i>et al.</i> (1955)	18 paires de soupe	Labo 1 paire/sujet Préférence	Domicile Panel différent 1 paire/sujet Préférence	Des divergences mais les résultats sont globalement identiques Produits très différents : résultats identiques Produits très proches : résultats différents (laboratoire plus discriminant que domicile)
Murphy <i>et al.</i> (1958)	6 sardines	Labo Classement Degré d'acceptation	Domicile Panel différent 1 paire/sujet Préférence Degré de préférence	Résultats identiques même si les écarts d'appréciation sont légèrement plus élevés à domicile qu'en labo
Calvin et Sather (1959)	Différentes catégories de produits, 15 paires	Labo 1 paire/sujet Préférence Note hédonique	Domicile Panel différent 1 paire/sujet Préférence Note hédonique	Note hédonique : résultats identiques pour 11 paires sur 15 (mais notes à domicile souvent plus élevées qu'en labo). Préférence : résultats identiques pour 12 paires sur 15. Résultats différents : domicile plus discriminant que labo (mais classements identiques)
McDaniel et Sawyer (1981)	6 whiskies testés contre le standard 6 paires	Labo 1 paire/sujet Note hédonique	Domicile Panel différent Test par paires Note hédonique	Résultats différents : domicile plus discriminant que labo (mais classements identiques)
Shepherd et Griffiths (1987)	3 œufs	Labo Note hédonique par rapport à l'idéal (goût et couleur du jaune d'œuf)	Domicile Panel différent Note hédonique par rapport à l'idéal (goût et couleur du jaune d'œuf)	Note du goût : résultats identiques Note du jaune d'œuf : résultats différents (notes à domicile supérieures aux notes en labo mais classement identique)
Hellemann <i>et al.</i> (1992)	1 repas complet	Labo Note hédonique	Domicile et cafétéria Panels différents Note hédonique	Résultats différents : notes en labo supérieures aux notes à domicile et à la cafétéria
Daillant-Spinnler et Issanchou (1995)	3 fromages blancs de matière grasse croissante	Labo Note hédonique par rapport à l'idéal	Domicile Panel identique Note hédonique par rapport à l'idéal	Produits à faible matière grasse : résultats identiques Produits à forte matière grasse : notes en labo supérieures aux notes à domicile
McEwan (1997)	10 biscuits salés aromatisés	Labo et salle (2 panels) Notes hédoniques (apparence, goût et texture)	Domicile Panel différent Notes hédoniques (apparence, goût et texture)	Résultats différents : - le produit le plus apprécié change - différences de notes entre les panels mais un panel ne note pas en moyenne plus fort qu'un autre.

Tableau 1 (suite) : Résultats des études ayant comparé des mesures hédoniques obtenues en conditions artificielles ou naturelles de consommation

Etude	Produits testés	Protocole en situation artificielle	Protocole en situation naturelle	Principales conclusions concernant la comparaison des méthodes
Meiselman <i>et al.</i> (2000)	Plat préparé au poulet	Labo Note hédonique globale et partielle	Restaurant (plat imposé) et cafétéria Panels différents Note hédonique globale et partielle	Résultats différents : les notes au restaurant sont supérieures aux notes en labo qui sont supérieures aux notes à la cafétéria
Pound <i>et al.</i> (2000)	3 barres chocolatées	Labo et salle (panel identique) Note hédonique	Domicile Panel identique Note hédonique	Résultats identiques : les notes à domicile et en labo sont supérieures aux notes en salle mais ce n'est pas significatif.
Posri <i>et al.</i> (2001)	4 thés	Salle Note hédonique	Domicile (long terme) Panel différent Note hédonique	Résultats différents : classement modifié
Hersleth <i>et al.</i> (2003)	8 vins	Labo Note hédonique	Salle de réception Panel différent Note hédonique	Résultats différents : les notes dans la salle de réception sont supérieures aux notes en labo
Kozłowska <i>et al.</i> (2003)	5 jus de pomme	Labo Note hédonique	Domicile Panel identique Note hédonique	Résultats différents pour un produit : le jus de fruits sans sucre ajouté est mieux noté à domicile qu'en labo.
King <i>et al.</i> (2004)	Thé, pizza et salade (2 échantillons par produit)	Salle Note hédonique	Restaurant Panel différent Note hédonique	Thé : les notes au restaurant sont supérieures aux notes en salle ; restaurant plus discriminant que salle Pizza : résultats identiques Salade : les notes au restaurant sont supérieures aux notes en salle ; salle plus discriminant que restaurant
De Graaf <i>et al.</i> (2005)	Multiples produits (plats, snacks et composants de repas)	Labo Note hédonique	Terrain : manœuvre militaire Panel différent Note hédonique	Produits de type <i>snacking</i> : résultats identiques Plats ou composants de repas : résultats différents avec une modification du classement
Hersleth <i>et al.</i> (2005)	6 fromages	Labo Note hédonique	Domicile Panel identique Note hédonique	Résultats identiques : classement des produits identique mais le domicile est légèrement plus discriminant que labo.
King <i>et al.</i> (2007)	Lasagnes, cannellonis, pain, thé glacé, salade	Salle Note hédonique	Restaurant Panel différent Note hédonique	Résultats différents : les notes au restaurant sont supérieures aux notes en salle
Petit <i>et al.</i> (2007)	2 thé glacés	Labo Note hédonique	Salle de réception Panel différent Note hédonique	Résultats identiques : les deux produits sont bien discriminés dans les deux tests

I.2 - UNE MESURE HEDONIQUE SENSIBLE AU CARACTERE ARTIFICIEL DES CONDITIONS DE DEGUSTATION

Ce paragraphe fait le point sur les travaux ayant abordé l'influence des conditions globales de dégustation lors d'un test hédonique. Nous nous intéressons ici plus particulièrement à la comparaison de données hédoniques obtenues en conditions de test contrôlées (en laboratoire ou en salle) que nous qualifierons d'artificielles avec les données hédoniques obtenues dans des environnements plus naturels de dégustation (domicile ou restaurant-test). Cette revue bibliographique a été réalisée pour répondre aux deux questions suivantes :

- Est-ce que les conditions de test artificielles aboutissent systématiquement à des résultats différents de ceux obtenus en conditions de dégustation naturelles ?
- Quelle est la nature de l'impact des conditions d'évaluation dans le cas de l'obtention de résultats différents ?

Le tableau 1 récapitule les méthodologies et résultats des 18 études récoltées qui se sont intéressées à la comparaison de tests consommateurs en situations artificielles ou naturelles.

1.2.1 - Les conditions de dégustation artificielles n'aboutissent pas systématiquement à des résultats différents de ceux obtenus dans des conditions naturelles

Une rapide observation des conclusions de l'ensemble des études ne montre pas de réel consensus en ce qui concerne l'existence ou non d'une influence des conditions de dégustation sur les résultats obtenus.

Certains auteurs, n'observent aucune différence de résultat entre les réponses hédoniques collectées en conditions artificielles ou naturelles de dégustation (Hersleth *et al.*, 2005 ; Murphy *et al.*, 1958 ; Petit *et al.*, 2007 ; Pound *et al.*, 2000 ; Shepherd *et Griffiths*, 1987).

D'autres études aboutissent à des conclusions plus mitigées concernant l'influence des conditions de dégustation. Ces études observent des différences de résultat selon les conditions de test uniquement pour certains échantillons (Daillant-Spinnler *et Issanchou*, 1995 ; Kozłowska *et al.*, 2003) ou uniquement pour certaines catégories de produits testées (Calvin *et Sather*, 1959 ; De Graaf *et al.*, 2005 ; King *et al.*, 2004 ; Miller *et al.*, 1955).

Enfin, certains travaux ont observé une modification systématique des résultats selon la méthodologie employée (Hellemann *et al.*, 1992 ; Hersleth *et al.*, 2003 ; King *et al.*, 2007 ; McDaniel *et Sawyer*, 1981 ; McEwan, 1997 ; Meiselman *et al.*, 2000 ; Posri *et al.*, 2001).

1.2.2 - Nature de l'impact des conditions de dégustation sur les résultats

Comme nous venons de le préciser, de nombreux travaux ont observé des modifications de résultat selon la méthodologie de test employée. Différents niveaux de modifications des résultats ont pu être constatés et il est important de les distinguer car ils illustrent le type de conséquences que peut avoir la mise en place de tests consommateurs en conditions de dégustation artificielles (en laboratoire ou en salle) sur l'information hédonique recueillie et donc sur les décisions stratégiques qui en découlent.

Trois niveaux de modifications ont été observés :

- une modification du niveau de notation de l'appréciation des aliments,
- une modification du niveau de discrimination des différents aliments comparés,
- une modification du classement hédonique des différents aliments comparés.

Certaines études ont observé uniquement une modification du niveau de notation selon les conditions de test. Ainsi, pour la majorité d'entre elles, les notes obtenues en conditions naturelles de consommation sont plus élevées que les notes obtenues en conditions artificielles (De Graaf *et al.*, 2005 ; Hersleth *et al.*, 2003 ; King *et al.*, 2004, 2007 ; Kozłowska *et al.*, 2003 ; Meiselman *et al.*, 2000 ; Pound *et al.*, 2000 ; Shepherd et Griffiths, 1987). Cependant, le phénomène inverse a été observé dans quelques cas (Daillant-Spinnler et Issanchou, 1995 ; Hellemann *et al.*, 1992).

En ce qui concerne le niveau de discrimination des produits testés, il ne semble pas y avoir de consensus quant à la méthode la plus discriminante. Alors que Miller *et al.* (1955) et Vickers *et al.* (1999) trouvent que les conditions de test artificielles aboutissent à des résultats plus discriminants que les conditions de consommation naturelles, Calvin et Sather (1959), Hersleth *et al.* (2005) et McDaniel et Sawyer (1981) observent le résultat inverse. King *et al.* (2004), quant à eux, sont confrontés aux deux types de résultats dans la même étude selon la catégorie de produits testée (les thés sont mieux discriminés au restaurant et les salades sont mieux discriminées en salle).

Enfin, les résultats dont les conséquences sont les plus importantes en termes de décision stratégique concernent les travaux qui ont mis en évidence une modification du classement hédonique des produits selon les conditions de test utilisées (De Graaf *et al.*, 2005 ; McEwan, 1997 ; Posri *et al.*, 2001).

1.2.3 - Est-il possible d'anticiper les conséquences de la mise en place d'un test en conditions artificielles ou naturelles de dégustation sur la mesure hédonique ?

Suite à cette revue bibliographique sur la comparaison des données hédoniques obtenues en conditions de consommation artificielles ou naturelles, il est difficile de pouvoir anticiper dans le cadre d'autres produits ou d'autres problématiques de test quelles seront les conséquences de l'utilisation d'un test en salle à la place d'un test à domicile sur la mesure hédonique obtenue.

Les résultats concernant l'impact ou non des conditions de test sur les mesures hédoniques sont très contrastés selon les études ce qui pourrait être attribué au fait que les méthodologies employées pour représenter les situations artificielles ou naturelles ne sont jamais strictement identiques d'une étude à l'autre (effectif, panels identiques ou non, protocoles mis en place..). Cependant, cette différence de résultats est également provoquée par le fait que chacune des ces études fait référence à des produits différents. En effet, les auteurs n'ayant pas observé d'influence des conditions de test sur leurs résultats l'interprètent par un usage plutôt nomade des produits testés ou encore une forte familiarité des sujets interviewés avec les produits. Par ailleurs, si nous nous concentrons sur les travaux qui ont réalisé plusieurs comparaisons basées sur des catégories de produits différentes, nous observons fréquemment qu'au sein d'une même étude, les résultats des comparaisons méthodologiques ne sont pas les mêmes bien que les procédures expérimentales utilisées soient strictement identiques. Ainsi, De Graaf *et al.* (2005) ont par exemple observé une influence très forte des conditions de test sur les produits consommés habituellement en repas et aucune influence sur les produits consommés plutôt en *snacking*. King *et al.* (2004) ont de leur côté trouvé un impact des conditions de test sur les thés et les salades mais pas pour les pizzas. Enfin, même au sein d'une même catégorie de produits comme les jus de pommes, Kozłowska *et al.* (2003) observent un impact des conditions de test sur la note hédonique d'un seul échantillon.

II - IDENTIFICATION DES VARIABLES DE CONTEXTE SUSCEPTIBLES D'INFLUENCER LA MESURE HEDONIQUE

Comme nous l'avons précédemment présenté, le jugement hédonique que nous mesurons lors des tests consommateurs est sensible à de nombreuses variables autres que les caractéristiques organoleptiques de l'aliment. Nous avons donc choisi de définir le contexte comme regroupant l'ensemble des variables qui entrent en jeu dans la formation de la perception hédonique et qui ne concernent pas les propriétés sensorielles de l'aliment.

Alors que l'influence du contexte sur la perception hédonique des aliments est maintenant admise, nous assistons ces dernières années à un réel désir de mieux le comprendre. Le paragraphe qui suit a pour objectif de répertorier les variables de contexte qui entrent en jeu lors d'une mesure hédonique et de faire le point sur les travaux qui ont exploré l'influence de ces facteurs contextuels.

De nombreuses études ont abordé l'influence individuelle de ces informations contextuelles activées par un sujet lors de la construction de son jugement hédonique en réponse à un aliment. Les études dont nous allons discuter ont été mises en place soit dans des conditions contrôlées soit lors de situations plus naturelles de consommation. Elles seront présentées dans ce chapitre sans distinction bien que les études réalisées en situation naturelle de consommation nous empêchent d'être certains que les autres variables de contexte restent constantes pendant la manipulation de la variable d'intérêt.

Nous verrons que la majorité des études ayant exploré l'influence des variables contextuelles se cantonnent à la manipulation des variables externes (visibles) sur lesquelles l'expérimentateur a une action. Cependant, nous aborderons également de manière plus conceptuelle d'autres informations contextuelles qui influencent fortement le jugement hédonique. Ces variables sont étroitement liées à l'individu et sont donc souvent difficiles à paramétrer lors de la mise en place d'un test consommateurs.

II.1 - INFLUENCE DES FACTEURS CONTEXTUELS EXTERNES

Alors que l'influence de certaines variables contextuelles se manifeste à la fois en situation de test et lors d'un acte naturel de consommation, d'autres variables sont plus spécifiquement liées aux conditions de dégustation dans le cadre des tests. Nous présenterons l'étude de ces deux types de variables et nous verrons comment elles sont prises en compte dans la mise en place des tests consommateurs et dans l'interprétation des données.

II.1.1 - La quantité d'aliment consommée lors de la dégustation

L'influence de la quantité proposée et consommée lors des tests hédoniques a été observée dans de nombreux tests en laboratoire. Ainsi, de nombreuses études ont montré que des fortes concentrations en sucre ou en sel sont plus appréciées lors de tests brefs que lors de tests de consommation ad libitum (Bellisle *et al.*, 1988 ; Lucas et Bellisle, 1987 ; Monneuse *et al.*, 1991a ; Perez *et al.*, 1994 ; Zandstra *et al.*, 1999). Cependant d'autres études donnent des résultats contradictoires (Dailliant et Issanchou, 1991 ; Popper *et al.*, 1989 ; Shepherd *et al.*, 1991). Par ailleurs, Hellemann et Tuorila (1991), Lähteenmäki et Tuorila (1994, 1995) et Popper *et al.* (1989) ont observé que des notes hédoniques obtenues après une consommation ad libitum sont plus élevées que des notes hédoniques obtenues après consommation d'une faible quantité.

Hellemann et Tuorila (1991) et Tuorila *et al.* (1994) concluent leurs travaux en précisant que des jugements hédoniques construits sur la consommation d'une dose habituelle de l'aliment sont plus pertinents que des jugements basés sur une brève exposition. Bien que les bonnes pratiques des tests consommateurs en salle précisent que la dose d'aliment présentée et ingérée doit correspondre à une dose habituelle du produit (AFNOR, 2000 ; Kôster, 1998), l'application de ce principe est rarement réalisée. En effet, cette pratique est généralement compromise par le fait que plusieurs échantillons de produits sont souvent testés dans la même séance et que les participants ne peuvent donc pas ingérer des doses réalistes de tous les échantillons. Les tests à domicile, quant à eux, permettent la consommation d'une quantité réaliste de produit à chaque occasion de consommation.

II.1.2 - La préparation de l'aliment testé

L'impact de la préparation de l'aliment sur son appréciation a été peu exploré. Par ailleurs, le plus grand nombre de travaux concerne uniquement la mise en évidence d'une amélioration de l'appréciation des aliments testés lorsque ces derniers sont présentés à une température culturellement appropriée (Boulze *et al.*, 1983 ; Cardello et Maller, 1982 ; Ryyanen *et al.*, 2001 ; Zellner *et al.*, 1988). Ces résultats obtenus pour des soupes ou des glaces ne sont pas très surprenants et la présentation des aliments à une température appropriée (soupe chaude et glace froide) est aujourd'hui bien intégrée dans les protocoles de test en salle. De manière moins radicale, la température de consommation d'un aliment peut être différente d'un individu à l'autre. C'est le cas notamment de certaines boissons qui peuvent être consommées à température ambiante ou fraîche. La prise en compte de la température habituelle de consommation peut dans ce cas-là devenir plus compliquée étant

donné que cela nécessite alors d'adapter la température de présentation aux pratiques des différents individus.

Deux études ont tenté de mesurer l'impact d'une préparation individuelle des produits testés sur leur appréciation. Matuszewska *et al.* (1997) ont comparé trois procédures de test pour l'évaluation d'échantillons de margarine. Premièrement, les sujets devaient étaler chaque échantillon sur des tranches de pain de la même façon qu'ils le font à domicile. La seconde méthode consistait à déguster des tranches de pain déjà tartinées avec 4 g de margarine. La troisième procédure représentait une situation dans laquelle les échantillons de margarine étaient dégustés seuls sans pain. Les résultats montrent que les différents échantillons sont plus appréciés et mieux discriminés lorsqu'ils sont consommés selon un protocole de préparation individuel. Posri *et al.* (2001) ont également comparé l'appréciation de plusieurs échantillons de thé mesurée selon trois conditions de préparation : une préparation imposée (tous les échantillons sont préparés avec les mêmes doses de lait et de sucre), une préparation individuelle mais contrôlée (les sujets choisissent leurs doses préférées de lait et de sucre et tous les échantillons sont présentés avec ces doses) et une préparation totalement libre (les sujets sont libres de doser les échantillons comme ils le souhaitent). Les auteurs observent encore une fois une forte influence de la préparation des produits sur leur appréciation. Par ailleurs, la comparaison des résultats obtenus selon ces trois protocoles de dégustation avec ceux d'un test à domicile montre que la préparation individuelle mais contrôlée des produits est le protocole qui permet le mieux de se rapprocher des données du test à domicile.

Cette variable contextuelle de préparation individuelle de l'aliment entre donc clairement en jeu dans la formation du jugement hédonique et est malheureusement peu prise en compte lors de la mise en place des protocoles de test consommateurs en conditions standardisées. On peut alors se poser la question de l'intérêt de récolter une réponse hédonique face à un aliment préparé de manière standardisée. Il peut être difficile d'interpréter par exemple une note d'appréciation donnée à un yaourt nature sans sucre par un individu qui les consomme généralement sucrés. En effet, un rejet des échantillons pourrait uniquement traduire le fait que l'individu n'aime pas les yaourts non sucrés alors que ceux-ci auraient été plus appréciés s'ils avaient été sucrés.

II.1.3 - L'environnement alimentaire

Outre la prise en compte de la préparation de l'aliment, l'impact de l'environnement alimentaire entourant la consommation de l'aliment testé est également souvent négligé dans la mise en place des tests consommateurs.

L'impact de l'appréciation d'aliments individuels sur l'appréciation globale d'un repas a été largement étudié. Alors que Hedderley et Meiselman (1995) et Turner et Collison (1988) concluent que tous les composants d'un repas ne participent pas de la même façon à l'appréciation globale du repas, Popper *et al.* (1989) et Tuorila *et al.* (1990) observent le contraire.

Cependant, nous nous intéressons plutôt à l'évolution de l'appréciation d'un aliment consommé seul ou au milieu d'autres aliments (un repas par exemple). Même si dès 1959, Eindhoven et Peryam observaient cet impact de l'environnement alimentaire sur l'appréciation d'un aliment, ce thème de recherche a fait l'objet de moins d'investigations. King *et al.* (2004) obtiennent une meilleure discrimination de deux pizzas quand elles sont consommées au milieu d'un repas que si elles sont dégustées de manière individuelle comme dans un test en salle classique. Dans de récents travaux confirmatoires, King *et al.* (2007) soulignent encore une fois l'importance d'intégrer la variable « repas » lors de la mise en place de protocoles de tests hédoniques en salle sur des produits habituellement consommés en repas ce qui est, à ma connaissance, rarement réalisé.

II.1.4 - L'information disponible sur l'aliment testé

Ce travail de thèse se situe dans le cadre des études consommateurs avec présentation des produits en aveugle. Cependant, le masquage des emballages n'étant pas systématiquement parfait (principalement lors des tests à domicile), nous ne pouvons pas exclure le fait qu'il arrive que certains participants reconnaissent les produits testés. Il nous a donc paru important de rappeler que les informations véhiculées par un emballage ou une marque influencent la formation du jugement. La modification de l'appréciation des produits due à leur étiquette a été largement étudiée dans les cas où l'information est d'ordre nutritionnel (Engell *et al.*, 1998 ; Kahkonen *et al.*, 1997) ou liée à des bénéfices sur la santé (Tuorila et Cardello, 2002). D'autres types d'informations ont également été explorés tels que le prix (Helgesen *et al.*, 1998 ; Moskowitz, 1995), la description du produit (Tuorila *et al.*, 1998), la cible de consommateurs potentiels (Jaeger, 2000), l'opinion des autres consommateurs (Cardello et Sawyer, 1992), l'étiquette de l'emballage (Lange *et al.*, 2002), le goût (Kahkonen *et al.*, 1997) ou encore l'origine des produits testés (Tuorila *et al.*, 1998). Le

rôle des attentes véhiculées par ces informations est bien évidemment majeur dans leur influence sur l'appréciation (pour une revue du rôle des attentes activées lors de la présentation d'information voir Lange *et al.*, 2000).

II.1.5 - Le moment de consommation

Nous avons trouvé seulement trois études ayant étudié spécifiquement l'impact de l'à-propos du moment de consommation sur les données hédoniques.

Birch *et al.* (1984) ont observé que des composants d'un petit déjeuner sont plus appréciés le matin que l'après-midi et que des composants de dîner sont plus appréciés l'après-midi que le matin. Cependant, la même étude réalisée par Kramer *et al.* (1992) n'a pas mis en évidence cette influence du moment de consommation pour les mêmes produits.

Cardello *et al.* (2000) montrent que l'effet du moment de la journée dépend du type de produit. Certains produits sont clairement associés à un moment de consommation (pizza) alors que d'autres peuvent être plus facilement consommés tout au long de la journée (céréales). Outre le fait qu'un sujet ait naturellement plus envie de consommer une pizza entre midi et 14h qu'à 10h du matin, l'influence du moment de la journée se manifesterait également par l'état physiologique des participants qui sont alors plus enclins à accepter un aliment s'ils ont faim que s'ils n'ont pas faim (cette variable contextuelle sera abordée ultérieurement).

La norme XP V09-500 (AFNOR, 2000) ainsi que d'autres recueils méthodologiques (Köster, 1998) préconisent la réalisation des tests de dégustation aux heures appropriées à la consommation du produit. Étrangement, alors que les tests consommateurs traditionnels s'attachent à imposer les mêmes conditions de dégustation pour tous les sujets afin que la variabilité des réponses hédoniques provienne uniquement des caractéristiques sensorielles de l'aliment dégusté, le moment de la dégustation est, lui, rarement standardisé. En effet, il est généralement impossible pour les instituts de louer une salle de test dans un lieu public pour seulement une ou deux heures. Ainsi, les séances de dégustation en salle sont souvent réparties sur l'ensemble de la journée pendant les horaires de passage des consommateurs dans ces lieux publics. Lors de tests sur des produits ayant un moment spécifique de consommation, il serait donc intéressant de distinguer les données récoltées aux heures appropriées à la consommation de celles récoltées lors de moments moins appropriés.

II.1.6 - L'environnement social

L'impact de l'environnement social a principalement été étudié sur la quantité d'aliment consommée. Roth *et al.* (2001) précisent que la quantité consommée par une personne est certainement guidée par le comportement alimentaire de son entourage immédiat. Plusieurs travaux ont mis en évidence un effet de facilitation sociale signifiant que les individus mangent plus quand l'acte de consommation a lieu avec d'autres personnes que quand ils sont seuls (Berry *et al.*, 1985 ; De Castro, 1990). Cependant, d'autres études obtiennent le phénomène inverse (Cardello *et al.*, 2000). Par ailleurs, les études de Feunekes *et al.* (1995) et de Pliner *et al.* (2003) ne révèlent pas d'effet direct du nombre de personnes entourant le sujet interviewé sur sa consommation. Ces auteurs suggèrent que les effets de facilitation sociale sont dus au temps passé à manger qui est plus long quand un individu mange accompagné que quand il est seul. Afin d'approfondir l'étude de ce phénomène de facilitation sociale, Clendenen *et al.* (1994) montrent que la relation qui lie les protagonistes du moment de consommation est un facteur qui contribue de manière significative à l'effet de facilitation sociale. Ainsi, une consommation dans un environnement familial serait plus sensible à la manifestation d'un effet de facilitation sociale qu'une consommation dans un environnement amical ou encore au milieu d'étrangers.

Etant donné que la présence d'autres personnes influencerait la quantité consommée et que la quantité consommée influencerait l'appréciation des produits, nous pouvons faire l'hypothèse que l'environnement social influence le niveau d'appréciation des aliments. Cependant, peu d'études ont exploré directement ce lien entre l'environnement social et les scores d'appréciation. L'étude de King *et al.* (2004) montre que l'appréciation d'une pizza diminue significativement quand elle est consommée au sein d'un groupe ce qui traduirait donc l'opposé du phénomène de facilitation sociale. Dans une autre étude, King *et al.* (2007) n'observent pas d'effet significatif d'un environnement agréable et social sur la notation de produits.

D'après Rozin (1996), la grande majorité des repas sont consommés en compagnie de quelqu'un. Nous pouvons donc nous interroger sur la pertinence des tests qui s'attachent à empêcher tout contact entre les différents interviewés. Cependant, le fait d'imposer un contact social entre des personnes qui ne se connaissent pas ne serait pas forcément une solution pertinente. En effet, Pliner *et al.* (2003) dans leur étude montrent que la facilitation sociale n'a un effet positif que dans le cas de groupes formés naturellement. Par ailleurs, des interviewés n'ayant pas forcément d'autres sujets de discussion pourraient avoir tendance à partager de manière non naturelle leurs opinions concernant les aliments testés. Ce partage d'opinions

pourrait alors modifier le jugement de personnes influençables. En effet, Cardello et Sawyer (1992) ont montré que l'appréciation d'un aliment peut être influencée par l'information donnée sur ce qu'en pensent d'autres personnes. Ainsi, la mise en place d'un test consommateurs en salle dans un environnement social nécessiterait de passer par une phase de pré-recrutement coûteuse afin de rassembler des personnes qui se connaissent.

II.1.7 - La possibilité de choisir l'aliment dégusté

Peu de travaux ont étudié spécifiquement la relation entre le degré de liberté dans le choix de l'aliment consommé et son appréciation.

Quelques études réalisées sur la monotonie ou la lassitude permettent d'appréhender l'impact de cette variable de choix sur l'appréciation d'aliments consommés de manière régulière. Zandstra *et al.* (2000) ont observé l'évolution de la note d'appréciation de trois échantillons obtenue après plusieurs dégustations à domicile selon différents degrés de choix. Ainsi, la consommation répétée du même échantillon pendant 10 semaines aboutit à une diminution des notes d'appréciation due à un effet de lassitude. En revanche, les scores d'appréciation des échantillons donnés par des personnes qui étaient libres de choisir les échantillons à consommer sur cette période ne subissent pas cette diminution. Kramer *et al.* (2001) ont également observé que le fait de laisser le choix de leur repas à des militaires sur le terrain permet de diminuer globalement l'effet de lassitude obtenu pour la dégustation imposée des mêmes aliments en laboratoire. La possibilité de choisir l'aliment que l'on souhaite consommer stimulerait également la quantité de produit consommée (Beatty, 1982 ; Zandstra *et al.*, 2000).

De Graaf *et al.* (2005) ont étudié ponctuellement l'impact du choix de l'aliment sur son appréciation. Ils ont obtenu en laboratoire des scores hédoniques moins élevés lors d'une dégustation imposée de tous les échantillons que lors d'une dégustation de certains échantillons choisis préalablement parmi une liste. De la même façon, King *et al.* (2004) ont testé l'impact de l'ajout d'une composante de choix du menu dans une procédure de test réalisée en salle en situation sociale de repas. Ils n'observent aucune influence de cette variable de choix sur l'appréciation globale du repas même si la note hédonique de certains aliments augmente quand ils ont été choisis.

D'après Rozin et Markwith (1991), la plupart des situations naturelles de consommation présentent des éléments de choix. Ainsi, nous pouvons supposer que l'absence de cette composante de choix dans les tests en conditions standardisées est une des causes de leur mauvaise validité prédictive. En ce qui concerne les tests à domicile, le sujet a le choix de

consommer le produit quand il le désire mais il a quand même pour obligation de consommer le produit testé et non un autre substitut. Dans une perspective d'amélioration de la validité prédictive des tests hédoniques, De Graaf *et al.* (2005) ont comparé les performances de données d'un test traditionnel avec celles d'un test présentant une possibilité de choix, pour prédire les notes hédoniques obtenues dans une situation naturelle de consommation. Ils ont ainsi observé que le test en laboratoire avec une possibilité de choix permet de mieux prédire les données obtenues en situation naturelle que le test en laboratoire sans possibilité de choix. Cependant, encore une fois, la prise en compte de cette variable de choix dans les protocoles de test hédonique impliquerait un budget supplémentaire lié à la participation d'un grand nombre de personnes afin que tous les échantillons testés soient choisis et évalués le même nombre de fois.

Pliner (1992) est en désaccord avec la critique de Meiselman (1992) concernant le fait que l'acte de consommation forcé pendant une séance de test en salle est très éloigné de l'acte de consommation dans une situation naturelle. Elle précise que de nombreuses situations naturelles présentent également des éléments proches de la situation imposée lors des tests. Elle donne comme exemple le contexte familial ou amical dans lequel, la plupart du temps, le repas est déposé sur la table sans l'accord de tous les participants. Ainsi, Pliner (1992) pose la question de savoir si un membre de la famille ou un invité peut raisonnablement rejeter l'aliment présenté ou dire qu'il veut manger autre chose. Le fait que les situations de consommation soient réalistes n'implique donc pas nécessairement une absence totale de contraintes. Le débat sur la nécessité d'intégrer une composante de choix dans les protocoles de test reste donc ouvert.

II.1.8 - L'effort physique nécessaire pour obtenir l'aliment

De manière générale, le jugement hédonique d'un aliment est le fruit d'un arbitrage entre l'effort, au sens large, fourni par le consommateur pour accéder à l'aliment et le bénéfice qu'il en retire. L'arbitrage entre l'argent dépensé et l'appréciation de l'aliment a été notamment étudié par Cabanac (1995). Par ailleurs, les méthodes d'analyse conjointe (Green, 1978) ou de choix sous contraintes économiques (Dransfield *et al.*, 1998 ; Lange *et al.*, 1999 ; 2000) permettent d'étudier plus spécifiquement les processus d'arbitrages mis en jeu lors de décisions d'achat. Cependant ce travail de thèse se situant dans le cadre de l'évaluation hédonique sans contrainte économique, nous nous intéresserons ici uniquement à la composante physique de l'effort fourni par le consommateur pour accéder à l'aliment. Cette caractéristique peut sembler dérisoire quant à son influence sur les mesures hédoniques.

Cependant, certains travaux ont montré qu'elle joue fortement sur le comportement de consommation des individus (Engell *et al.*, 1996 ; Levitz, 1975 ; Meyers *et al.*, 1980). Bien que ces études n'aient pas observé d'impact sur les notes hédoniques, elles mettent en évidence un réel impact de l'accessibilité des produits sur le choix ou sur la quantité consommée. Par exemple, Engell *et al.* (1996) ont observé que les sujets boivent plus d'eau quand celle-ci est posée sur la table que quand ils doivent se lever pour aller en chercher. Meiselman *et al.* (1994) ont réalisé le même type d'étude dans une cafétéria étudiante dans laquelle ils ont changé de place la distribution de friandises. La première semaine les friandises sont présentées sur le lieu de paiement, à la fin du comptoir de service et la deuxième semaine elles sont présentées sur un autre comptoir de service et nécessitent un nouveau paiement. Les auteurs observent alors une chute de la sélection des friandises la deuxième semaine ainsi qu'une diminution des notes hédoniques données par les étudiants qui les ont achetées. Le rôle des attentes hédoniques dans la diminution des notes hédoniques est probablement prépondérant. En effet, l'effort est fourni par l'individu uniquement s'il a extrêmement envie de manger la friandise. Ainsi, l'attente de ces individus est très élevée en comparaison à celle des individus de la semaine précédente et l'appréciation de la friandise pourrait donc subir un effet de contraste si elle ne comble pas ces attentes élevées. De la même façon, nous pouvons imaginer que nous sommes plus exigeants envers un plat pour lequel la préparation aura été longue et fastidieuse que pour le même plat réchauffé au four à micro-ondes.

Etant donné que lors des tests consommateurs les produits sont directement présentés au sujet, les conditions d'effort sont généralement inexistantes ce qui pourrait donc entraîner une sur-estimation de l'appréciation des produits en comparaison à une situation naturelle dans laquelle la consommation du même produit aurait nécessité un effort d'accessibilité ou de préparation.

Les variables de contexte présentées précédemment influencent naturellement le jugement hédonique d'un consommateur dans sa vie quotidienne et s'appliquent également aux situations de test. Nous allons à présent détailler deux variables de contexte supplémentaires dont l'influence se manifeste sur le jugement hédonique uniquement dans le cadre des situations de test puisqu'elles proviennent plus spécifiquement des protocoles de test.

II.1.9 - Le contexte évaluatif apporté par les autres produits testés

En analyse sensorielle, le contexte d'évaluation est historiquement défini comme l'ensemble des échantillons dégustés à partir du moment où les niveaux de notation (appréciation ou intensité d'attributs sensoriels) d'un produit peuvent varier en fonction des autres produits dégustés dans la même séance. Lors des tests hédoniques, cette variable contextuelle n'est pas systématiquement présente étant donné qu'il arrive que l'individu ne déguste qu'un seul échantillon. Cependant nous devons envisager son impact lors de l'évaluation de plusieurs échantillons.

Lawless (1983), McBride (1985) et Risky *et al.* (1979) ont montré que si un stimulus est dégusté au milieu d'une série de stimuli présentant de très bonnes qualités organoleptiques, il ne sera pas jugé de la même manière que s'il est évalué au milieu d'un ensemble de stimuli de qualité médiocre. Nous pouvons également trouver le même type d'influence quand un consommateur est déçu par un repas dégusté dans un restaurant qui propose habituellement des produits de qualité. Dans ce cas, le contexte comparatif n'est pas immédiat mais est réalisé par rapport à des expériences passées et mémorisées.

Outre le niveau de qualité de l'ensemble des stimuli testés, l'environnement évaluatif peut avoir un impact sur l'appréciation globale du produit selon la procédure de présentation des échantillons utilisée lors d'un test. En effet, l'appréciation d'un échantillon testé peut être influencée par l'échantillon testé avant ou par l'anticipation de l'échantillon qui suivra. Les études ayant comparé des évaluations réalisées en simultané ou en monadique pur permettent d'évaluer l'impact que peut avoir une évaluation comparative à la place d'une évaluation dans l'absolu. Ainsi, Bayton et Thomas (1954) et McBride (1986) ont observé des classements de produits identiques mais des résultats plus discriminants avec les procédures en simultané qu'avec les procédures en monadique pur. D'autres auteurs ont comparé les procédures en monadique séquentiel avec la procédure en simultané. Certains n'ont pas mis en évidence de différence de résultats (Lähteenmäki et Tuorila, 1995 ; McBride, 1986) alors que d'autres ont observé des niveaux de discrimination différents (Delarue *et al.*, 2001 ; Schwartz et Pratt, 1956) ou des inversions de classement (Lähteenmäki et Tuorila, 1994 ; Leon *et al.*, 1999). Par ailleurs, il semblerait que l'influence de la procédure de présentation est d'autant plus importante que les échantillons testés sont proches sensoriellement.

L'influence de l'environnement évaluatif apporté par les autres échantillons testés soulève la question de la représentativité des jugements hédoniques générés sur une base comparative. Le jugement le plus proche de la réalité correspondrait plutôt à une évaluation par rapport à la référence interne de l'individu (c'est-à-dire par rapport à ses attentes

concernant la catégorie de produits) qu'à une évaluation basée sur la comparaison de plusieurs produits imposés. Ainsi, la mise en place des procédures de présentation des produits en monadique pur serait plus pertinente. Cependant, le cadre des tests consommateurs étant généralement basé sur l'évaluation de plusieurs échantillons, les procédures en monadique séquentiel sont souvent privilégiées au détriment des procédures en monadique pur plus coûteuses. En ce qui concerne ces tests en monadique séquentiel, Peryam et Pilgrim (1957) parlent d'une autre variable de contexte qui réfère à la position des produits évalués. Ils observent que le premier échantillon présenté tend à être noté plus positivement que les autres échantillons. D'autres auteurs ayant observé ce phénomène de sur-notation du premier produit (Filipello, 1956 ; Simone et Pangborn, 1957) proposent l'utilisation d'un produit de chauffe dont l'évaluation n'est pas prise en compte dans l'analyse des données. Par ailleurs, outre l'effet sur le premier produit dégusté, l'ordre de présentation global de l'ensemble des échantillons a un impact sur les évaluations ce qui peut être contrecarré par des plans de présentation des échantillons équilibrés sur l'ensemble du panel.

II.1.10 - Le design du questionnaire

En psychologie et en sociologie, l'utilisation de questionnaires attitudinaux est fréquente et la littérature sur les méthodologies d'enquêtes par questionnaire présente de nombreuses études concernant l'impact du design du questionnaire sur les réponses. Les auteurs définissent donc le design du questionnaire comme une variable de contexte à part entière. Les composantes du design et de l'administration du questionnaire sont à elles seules sources de nombreux effets de contexte largement étudiés dans la littérature :

- le contenu des questions,
- le format des questions (ouvertes, fermées) et l'échelle de réponse proposée pour les questions fermées,
- le mode d'administration (face-à-face, téléphone, auto-administré),
- le nombre de questions (longueur du questionnaire),
- l'ordre des questions.

Comme le suggèrent Schuman et Presser (1981), l'ordre des questions est probablement la cause la plus importante (après les erreurs d'échantillonnage) de résultats d'enquêtes surprenants et non répétables. L'effet d'ordre des questions est défini comme les autres effets de contexte et signifie donc que les réponses obtenues quand les questions sont posées dans un certain ordre sont différentes de celles qui auraient été obtenues avec un ordre différent. Cet effet consiste en fait à l'activation d'informations qui peuvent être, soit contenues

directement dans la question précédente, soit activées par le sujet pour répondre à la question précédente (Strack, 1992). La mise en évidence des effets d'ordre des questions dans les enquêtes a été observée majoritairement dans la littérature de la psychologie. Cependant, la plupart des résultats de ces études peuvent avoir des implications dans le domaine des tests consommateurs et plus spécialement pour le débat concernant la présence de questions diagnostiques en plus de la question sur l'appréciation globale. En effet, bien que l'objectif des tests hédoniques soit généralement de récolter une information sur l'appréciation globale des produits, les expérimentateurs profitent de ces tests pour tenter d'obtenir des informations plus précises concernant les produits dégustés. L'ensemble de ces informations supplémentaires est collecté par l'intermédiaire de questions diagnostiques regroupant différents types de questions : des questions ouvertes sur les raisons d'appréciation ou de rejet, des questions concernant le niveau d'appréciation de certains attributs sensoriels et des questions concernant l'intensité de certains attributs sensoriels. Etant donné qu'il est impératif que l'évaluation globale fournie par un consommateur pour un ensemble de produits soit représentative de ses préférences réelles et ne soit pas biaisée par d'autres variables incluses dans le déroulement de la mesure, la présence de ces questions diagnostiques nous oblige à étudier l'impact qu'elles peuvent avoir sur l'appréciation globale donnée par le sujet aux produits. Earthy *et al.* (1997) et Popper *et al.* (2004) affirment que les questions diagnostiques entraînent un biais dans les résultats des tests consommateurs tels qu'ils sont mis en place le plus souvent c'est-à-dire avec une seule séance de dégustation sur des produits présentés de manière séquentielle. Les questions diagnostiques ont tendance à amener le sujet à réaliser son jugement global sur la base d'une analyse très précise des caractéristiques sensorielles abordées par les questions diagnostiques, chose qu'il n'aurait pas fait spontanément si le questionnaire ne l'y avait pas sensibilisé. Ainsi, même si l'appréciation globale est demandée pour chaque échantillon avant les questions diagnostiques, seule la construction du jugement global concernant le premier produit dégusté est épargnée par la sensibilisation aux caractéristiques sensorielles.

De nombreux travaux concernant la mesure de l'impact du design du questionnaire ont exploré l'influence d'évaluations spécifiques sur une évaluation globale. Différents types de questions peuvent accompagner l'appréciation globale dans un questionnaire. Il nous a donc paru intéressant d'aborder leur influence sur l'appréciation globale de manière séparée.

II.1.10.1 - Influence de l'analyse des raisons d'appréciation sur l'évaluation globale

Dans le domaine des tests consommateurs organoleptiques, nous avons trouvé une seule étude (Vickers *et al.*, 1993) ayant exploré l'impact de l'analyse des raisons de préférence sur l'évaluation globale des produits. Cependant, les auteurs n'observent pas d'influence de cette question ouverte sur l'appréciation globale des produits.

D'autres études dans les domaines du marketing et de la psychologie ont, quant à elles, mis en évidence l'existence d'un impact de l'analyse des raisons de préférence ou de rejet sous forme de questions ouvertes sur l'évaluation globale. Gross (1964) montre que le niveau d'intérêt mesuré pour un produit peut être significativement diminué par l'ajout de l'analyse des inconvénients du produit. De leur côté, Wilson et Schooler (1991) observent que la variabilité des notes individuelles des sujets diminue en présence de questions sur l'analyse des raisons de préférence envers des confitures. Ces résultats reflètent donc que ce type d'introspection modère l'évaluation globale des sujets qui finissent par avoir du mal à distinguer les produits. En effet, plus les différents produits sembleront avoir à la fois des avantages et des inconvénients, plus ils seront perçus comme similaires en termes de nombre de points positifs et négatifs et donc plus ils seront appréciés de la même façon. Une autre étude réalisée par Wilson et Schooler (1991) montre qu'un choix fait après analyse des raisons n'est pas optimal et donc biaisé. Les auteurs ont contacté des sujets quelques semaines après un test lors duquel ils devaient emporter leur poster préféré et ils leur ont demandé s'ils étaient satisfaits de leur choix. Les sujets ayant fait leur choix après en avoir analysé les raisons étaient significativement moins satisfaits que ceux ayant choisi spontanément. Ainsi, avec le temps, les sujets reviennent probablement aux schémas qu'ils ont l'habitude d'utiliser et s'ils font un choix sur la base d'un schéma de pondération différent, ils peuvent regretter ce choix. Les hypothèses de Wilson et Schooler (1991) concernant la perturbation des préférences due à l'analyse des raisons de leurs sentiments est la suivante. Les sujets ne savent généralement pas pourquoi ils ont tel ou tel sentiment envers un stimulus. Quand ils réfléchissent à propos des raisons, ils centrent leur raisonnement sur des explications saillantes et possibles. Mais, ce qui apparaît comme une cause possible et ce qui détermine réellement les réactions des sujets face à un stimulus n'est pas forcément identique.

Pour conclure, demander à un sujet de décrire les raisons de sa préférence ou de son rejet avant de déclarer son jugement global peut entraîner une moins bonne discrimination des produits ou, dans le pire des cas, une préférence éloignée de la réalité.

II.1.10.2 - Influence de la notation de l'intensité d'attributs sensoriels sur l'évaluation globale

Les études réalisées dans le domaine de l'analyse sensorielle (Mela, 1989 ; Popper *et al.*, 2004 ; Vickers *et al.*, 1993) sont consensuelles et n'observent pas d'influence de la notation de l'intensité d'attributs sensoriels sur l'appréciation globale. Les auteurs mettent cette absence d'effet sur le compte de la nature de la question. En effet, la question d'intensité des attributs sensoriels ne possède pas de composante affective contrairement aux autres types de questions. Nous pouvons cependant remettre en question cette explication étant donné que même si l'échelle proposée reste descriptive, certains attributs sensoriels véhiculent des attentes hédoniques culturellement établies comme le sucré qui est plutôt apprécié alors que l'amer est plutôt rejeté. Par ailleurs, quelle que soit l'échelle proposée, le seul fait de sensibiliser les sujets à certaines caractéristiques spécifiques des produits devrait rendre ces attributs plus saillants et donc influencer la construction du jugement global du produit.

II.1.10.3 - Influence des évaluations hédoniques partielles (appréciation spécifique et notation par rapport à l'idéal) sur l'évaluation globale

Il est intéressant de citer ici quelques études réalisées dans le domaine de la mesure de la Qualité de Vie (QoL) dans laquelle le sujet doit donner son niveau de satisfaction concernant sa vie en général et différents domaines de sa vie comme le mariage, le travail, la santé. Ainsi d'après l'étude d'Andrews et Withey (1976) ce sont par exemple les domaines de la famille et du travail qui, spontanément, influencent le plus la qualité de vie alors que ce sont les domaines de l'amitié et des études qui influencent majoritairement la qualité de vie une fois que tous les domaines ont été activés. D'après McClendon et O'Brien (1988), dans la pratique, les questionnaires de QoL présentent en général les questions spécifiques avant la question globale afin d'être certain que toutes les dimensions spécifiques ont été prises en compte dans le jugement global. Placer les questions concernant la satisfaction des domaines spécifiques avant la question globale assure que tous les domaines ont été récemment activés en mémoire avant de répondre à la question générale. Ainsi, l'influence de la notation de la satisfaction par domaine sur la notation de la qualité de vie globale, est bien connue et même recherchée dans les questionnaires sur la QoL.

Dans le domaine du marketing, Moorthy (1991) a comparé l'évaluation de la satisfaction globale d'un service obtenue spontanément avec celle obtenue après l'évaluation de la satisfaction des différents attributs du service. L'évaluation de la satisfaction globale

spontanée est significativement plus faible que l'évaluation de la satisfaction globale obtenue après celle des différents attributs.

Les études concernant le domaine de l'analyse sensorielle ont utilisé le format habituel des questions sur les attributs sensoriels : les échelles de notation par rapport à l'idéal dites JAR (« *Just About Right* ») ou les questions sur la notation du niveau d'appréciation partielle de certains attributs. Vickers *et al.* (1993) ne trouvent aucun impact de la notation de quelques appréciations partielles sur la notation hédonique globale. Cependant, Earthy *et al.* (1997) observent une influence des questions d'appréciation spécifique de type JAR sur l'évaluation globale de sept produits. L'ajout des questions spécifiques entraîne la diminution des notes des produits les plus appréciés et l'augmentation des notes des produits les moins appréciés. Nous retrouvons donc de nouveau l'impact des évaluations partielles sur la moins bonne discrimination des produits en termes d'appréciation globale. Enfin, une dernière étude plus récente de Popper *et al.* (2004) montre que l'effet des questions JAR sur l'appréciation globale est plus prononcé que celui apporté par les questions d'appréciation partielle. Bien que ces deux types de questions soient basées sur une composante affective, l'impact plus important des questions JAR proviendrait, d'après les auteurs, de la notion de « produit idéal » qui n'existe pas dans les questions d'appréciation partielle.

Bien que l'influence des questions diagnostiques sur les appréciations globales soit aujourd'hui reconnue et leur utilisation déconseillée, il est encore très difficile de convaincre les commanditaires des tests consommateurs de privilégier les questions d'opinion globale au détriment des questions plus spécifiques. Ainsi, lors du congrès Pangborn en 2005, MacFie propose à la communauté sensorielle de parler de « scores hédoniques modulés » dans un souci de rigueur scientifique lorsque l'appréciation globale est mesurée accompagnée de questions diagnostiques. De leur côté, Popper *et al.* (2004) proposent de réaliser deux dégustations différentes afin d'éviter une pollution de l'évaluation globale par les évaluations spécifiques. La première consisterait à faire noter uniquement l'appréciation globale de tous les produits les uns à la suite des autres. La seconde pourrait être utilisée pour faire évaluer les questions spécifiques sur chacun des produits. Cette solution est également celle que conseille la norme XP V09-500 (AFNOR, 2000). L'ensemble des études hédoniques menées dans ce travail de thèse a été mis en place selon ce design de protocole en deux phases distinctes de manière à collecter des réponses d'appréciation globale les plus proches de la réalité dans laquelle un consommateur analyse rarement de manière consciente les caractéristiques sensorielles des produits afin de se faire un jugement.

II.2 - INFLUENCE DES FACTEURS CONTEXTUELS INTERNES

Earthy *et al.* (1997) définissent le contexte comme « regroupant toutes les variables indésirables incluses dans le déroulement d'une étude qui influencent les résultats ». Cette définition traduit en elle-même l'impuissance des expérimentateurs face à l'influence des variables de contexte sur les jugements évaluatifs. De fait, la standardisation des conditions de test a pour objectif de contrôler l'influence des variables de contexte afin de ne mesurer que l'effet des caractéristiques organoleptiques du produit sur la réponse hédonique. Cependant, cette solution est envisageable uniquement pour les facteurs contextuels énoncés précédemment. En effet, d'autres variables indépendantes du protocole de dégustation ou des caractéristiques sensorielles de l'aliment entrent également en jeu dans la formation de la perception hédonique de l'aliment mesurée en test consommateurs. Nous nous devons donc de mentionner l'existence de ces variables contextuelles liées à l'individu et à son expérience du produit. Même si ces variables sont difficilement paramétrables lors d'un test consommateurs, leur détermination est importante et pourrait être prise en compte pour l'interprétation des données hédoniques.

Parmi les facteurs liés à l'individu qui influencent son jugement hédonique, nous avons identifié deux sources possibles de facteurs contextuels qui peuvent intervenir lors de l'émission d'un jugement hédonique face à un aliment :

- le vécu alimentaire de l'individu,
- l'état interne de l'individu (état physiologique et psychologique).

II.2.1 – Le vécu alimentaire de l'individu

II.2.1.1 - Le nombre de contacts entre l'individu et l'aliment

Parmi les facteurs liés au vécu alimentaire de l'individu, l'influence du nombre de contacts entre l'individu et l'aliment a été largement démontrée et théorisée. Ainsi, de nombreux auteurs ont observé que des expositions répétées peuvent entraîner une diminution de la sélection de l'aliment ou de son évaluation (effet de lassitude), que ce soit lors d'un test en salle (Meiselman *et al.*, 2000 ; Muller, 2004 ; Vickers et Holton, 1998) ou lors d'évaluations en situation naturelle de consommation (Hetherington *et al.*, 2000 ; Kramer *et al.*, 2001 ; Levy et Köster, 1999 ; Zandstra *et al.*, 2000 ; 2004). Dans ce sens, Stang (1975) déclare que l'on a absolument besoin de variété pour éviter une lassitude provenant de la familiarité. Cependant, le phénomène inverse a également été mis en évidence et formulé par

Zajonc (1968) comme modèle de simple exposition. Cette théorie considère que les individus préfèrent les objets familiers aux objets non familiers et que, plus ils sont exposés fréquemment aux objets, plus ils se mettront à les aimer. Ainsi, il semblerait que le phénomène d'amélioration de l'appréciation (simple exposition) aurait lieu pour des produits non familiers ou nouveaux alors que le phénomène de lassitude serait plutôt observé pour des produits peu complexes et familiers. Par ailleurs, d'après Dember et Earl (1957), chaque personne en fonction de l'expérience qu'il possède a un niveau de complexité optimal différent. Berlyne et Slater (1957) soulignent cependant pour modérer la théorie de Zajonc que des aliments « nouveaux » peuvent également attirer l'attention et permettre de rompre la monotonie. Cette différence de comportement entre la recherche ou le rejet de produits nouveaux fait appel à une caractéristique intrinsèque au sujet et concerne la néophobie et la néophilie.

Quel que soit l'impact des expositions répétées sur l'évolution des préférences, l'intégration de la notion de familiarité de l'aliment présenté pour le sujet fait donc encore une fois appel à son vécu alimentaire. L'importance de ce vécu concernant le produit dégusté en test est capitale dans l'interprétation des données. La prise en compte du statut d'utilisateur de la catégorie de produits ou même de la marque de produit testée (consommateurs potentiels, occasionnels et réguliers) dans l'interprétation des données hédoniques est donc importante. Par ailleurs, différents protocoles de test consommateurs ont vu le jour pour évaluer l'évolution de l'appréciation des produits suite à des consommations répétées (les tests de lassitude, les tests d'aversion).

II.2.1.2 - L'anticipation des conséquences post-ingestives de l'aliment

Une autre variable liée au vécu alimentaire de l'individu ayant comme conséquence le rejet ou le souhait de consommer un aliment concerne l'anticipation des conséquences physiologiques post-ingestives de l'aliment.

D'après Pavlov (1927), les réponses conditionnées se mettent en place à la suite de l'association entre une certaine stimulation céphalique aux caractéristiques particulières et certains effets nutritionnels qui suivent l'ingestion de l'aliment. Les caractéristiques sensorielles des aliments (aspect, texture, odeur et goût) sont dues à des composants qui, en général, ne sont pas nutritionnellement actifs (sauf pour les sucres et le sel). Cependant, la sensibilité intestinale fournit les informations sur les effets post-ingestifs des aliments ingérés, qu'ils soient bénéfiques (soulagement du déficit énergétique, soulagement d'une carence) ou nocifs (malaises intestinaux). D'après Guevara-Aguilar *et al.* (1988), les informations

sensorielles provenant de la sphère orale et les informations venant de la sphère viscérale convergent vers une même zone hypothalamique. Au fil des prises répétées d'un aliment, ces caractéristiques sensorielles sont progressivement associées aux effets post-ingestifs ressentis dans les heures qui suivent l'ingestion. Le traitement au niveau central de la trace mnésique de l'aliment associée à celle des effets post-ingestifs a pour effet de moduler la composante affective associée à l'image sensorielle du produit et ce processus permet alors la sélection alimentaire. Ce mécanisme d'apprentissage fait donc, une fois de plus, référence au vécu alimentaire du sujet et peut être extrêmement puissant et entraîner des aversions alimentaires.

Le contexte qui va suivre l'acte alimentaire intervient également dans le jugement évaluatif et donc dans le comportement face à un aliment. En effet, d'après Rozin (1990), l'association entre les conséquences physiologiques de l'ingestion et les caractéristiques sensorielles de l'aliment va guider le comportement qui s'ajustera aux besoins présents et anticipés. Ces conséquences anticipées peuvent être soit plutôt stables et liées à des croyances (aliment santé, allégé, bio..), soit complètement dépendantes du contexte à venir (besoin d'énergie pour une activité sportive). Le risque d'interviewer lors d'un test consommateurs des sujets ayant une aversion forte pour la catégorie d'aliment testée est nul si le recrutement est réalisé correctement. Cependant, l'influence de l'anticipation d'effets post-ingestifs possibles existe et peut intervenir sur la quantité d'aliment consommée pendant le test et donc sur son appréciation. Par ailleurs, le fait que lors d'un test bref de dégustation les sujets n'aient pas le temps de ressentir les effets bénéfiques ou nocifs dus à l'ingestion du produit peut biaiser fortement la réponse hédonique.

II.2.2 - L'état interne de l'individu

La plupart des facteurs contextuels contrôlables listés précédemment jouent un rôle important dans l'état psychologique et physiologique des sujets durant une séance de test ce qui peut donc influencer la formation du jugement hédonique.

L'étude directe de l'influence de l'état interne des sujets (physiologique et psychologique) sur les évaluations hédoniques a donc été abordée dans quelques travaux.

II.2.2.1 - L'état physiologique de l'individu

L'état physiologique du sujet semble être un déterminant important du plaisir évoqué lors d'une consommation. En effet Cabanac *et al.* (1968) ont montré qu'une saveur sucrée peut être particulièrement appréciée à jeun et non appréciée en état de satiété. De la même façon, Laeng *et al.* (1993) ont mis en évidence que les sujets notent des boissons citronnées

sucrées comme étant moins plaisantes quand elles sont testées juste après un repas. Hill (1974) et Bell (1993) dans leurs travaux respectifs ont également mis en évidence un effet significatif de l'état de faim des sujets sur les préférences et les notes hédoniques aboutissant à une moins bonne discrimination des différents produits lorsque les interviewés ont faim. Brunstrom *et al.* (1997) ont, quant à eux, observé que le taux d'humidité de la bouche influence l'appréciation de différentes boissons.

Plus spécifiquement, il peut y avoir une évolution de l'appréciation des aliments tout au long de la dégustation d'un produit (Taylor et Pangborn, 1990) ou juste après la dégustation du produit lors d'une occasion de consommation. Les physiologistes ont mis en évidence deux phénomènes responsables de l'instabilité de la réponse hédonique à court terme : l'alliesthésie et le rassasiement sensoriel spécifique.

Dès 1971, Cabanac crée le mot d'*alliesthésie* pour désigner le changement de perception hédonique dû à la modification de l'état interne du sujet après la consommation du produit et non à la modification de la sensation. Deux formes d'alliesthésie sont identifiées : l'alliesthésie négative a lieu quand la modification de l'état interne entraîne une décroissance de l'appréciation alors que l'alliesthésie positive correspond à une évolution inverse. L'alliesthésie alimentaire négative peut être considérée comme étant un phénomène réflexe susceptible de créer un feedback pour freiner l'ingestion au bout de 15 ou 20 minutes (Cabanac *et al.*, 1973). On retrouve dans ce phénomène le problème des tests en salle qui, faute de temps et de quantités consommées suffisantes, ne permettent pas à ce mécanisme de se manifester.

Le phénomène de rassasiement sensoriel spécifique correspond à une diminution du plaisir sensoriel associé à un aliment en particulier et qui est induite par l'ingestion orale de ce seul aliment (Rolls, 1987 ; Rolls *et al.*, 1984). Le plaisir provoqué par d'autres aliments reste quant à lui intact. D'après Fantino (1992), ce rassasiement serait dépendant des qualités sensorielles de l'aliment consommé. Contrairement à l'alliesthésie, cette diminution du plaisir quasi-immédiate atteindrait son maximum 2 minutes après la mise en bouche et durerait environ 20 minutes. L'évolution de l'appréciation de l'aliment pendant la dégustation nous pousse alors à nous interroger sur le moment où l'on souhaite récupérer l'appréciation de l'aliment lors d'un test en salle : dès les premières bouchées c'est-à-dire à son maximum ou après que le rassasiement sensoriel ne se soit déclenché.

II.2.2.2 - L'état psychologique de l'individu

L'état psychologique du sujet pendant le test est également un facteur déterminant dans son jugement hédonique. Généralement, l'influence de l'état psychologique du sujet se traduit par des effets d'assimilation c'est-à-dire des jugements évaluatifs congruents avec l'affect de l'individu qui juge. De nombreuses études ont observé que l'amorçage d'un état d'humeur positif peut entraîner des attitudes plus positives que l'induction d'un état d'humeur neutre ou négatif (Bower, 1981 ; Deldin et Levin, 1986 ; Isen *et al.*, 1978). Cependant, cet effet n'a pas été au centre de beaucoup d'études dans le domaine des tests hédoniques. Siegel et Risvik (1987) ont amorcé différents états cognitifs en demandant à deux groupes de sujets de remplir deux questionnaires basés soit sur des phrases négatives soit sur des phrases positives. Les notes hédoniques données à l'aliment testé sont plus élevées pour le groupe amorcé positivement que pour le groupe amorcé négativement. Bonin *et al.* (2001) ont également étudié l'influence d'un amorçage positif ou négatif sur le choix d'aliments. Avant d'entrer dans une cafétéria, les sujets sont exposés à une phase de rappel autobiographique liée à une occasion positive ou négative de consommation de pâtisserie. Seulement 5 % des sujets amorcés négativement choisissent ensuite une pâtisserie alors que 33 % des sujets amorcés positivement en choisissent une. Ainsi, le simple fait de remémorer un évènement négatif permet de modifier le comportement alimentaire. Nous retrouvons ici l'influence du vécu alimentaire du sujet sur l'appréciation des produits.

Enfin, l'influence de la rémunération des sujets sur l'appréciation des aliments a été également soulignée par Bell (1993) qui montre qu'elle peut entraîner des notes plus élevées et diminuer la discrimination entre les échantillons. Cependant, le fait de ne pas rémunérer les sujets pose un problème quant à leur faible implication dans le test. Nous sommes donc encore une fois face à un problème délicat quant au choix d'indemniser ou non les participants.

L'état interne de l'individu au moment où il déguste l'aliment testé est donc primordial dans la formation de sa perception hédonique (état physiologique) ou dans l'émission de son attitude (état psychologique). Il semble illusoire de contrôler entièrement ces aspects lors de la mise en place des tests organoleptiques. Par ailleurs, cette diversité des états des individus s'exprime de manière naturelle dans la vie de tous les jours et ne nécessite donc pas une standardisation. Cependant, l'évaluation de l'état interne des individus au moment de l'évaluation des produits (l'état de faim, l'état de soif, l'état d'humeur) serait un outil précieux pour aider l'expérimentateur à interpréter la mesure hédonique.

III - BILAN : DISCUSSION SUR LA PERTINENCE DES REPONSES HEDONIQUES COLLECTEES EN CONDITIONS STANDARDISEES D'EVALUATION

L'examen de la littérature nous a permis de mettre en évidence la variabilité des jugements obtenus dans des conditions différentes de consommation, que celles-ci soient artificielles ou naturelles. La figure 9 récapitule l'ensemble des variables de contexte qui entre en jeu lors de l'obtention d'une mesure hédonique déclarative. Nous comprenons donc mieux pourquoi il n'est pas rare d'observer une incohérence entre des jugements émis par les mêmes personnes lors de différentes séances de dégustation pour un produit dont la composition n'a pas été modifiée. Köster (1981a) affirme que « lorsque des réponses hédoniques sont récoltées, les circonstances pourraient elles-mêmes influencer le jugement plus que les stimuli eux-mêmes ». Ce phénomène s'avère particulièrement problématique si l'on considère que l'objectif premier des tests consommateurs est de prédire les conduites futures des consommateurs et donc le devenir du produit sur le marché. Mais que dire à un chef de produit quant au positionnement probable de son produit sur le marché lorsque ce produit fait l'objet d'évaluations répétées différentes et qu'il n'est probablement pas responsable de cette variation ?

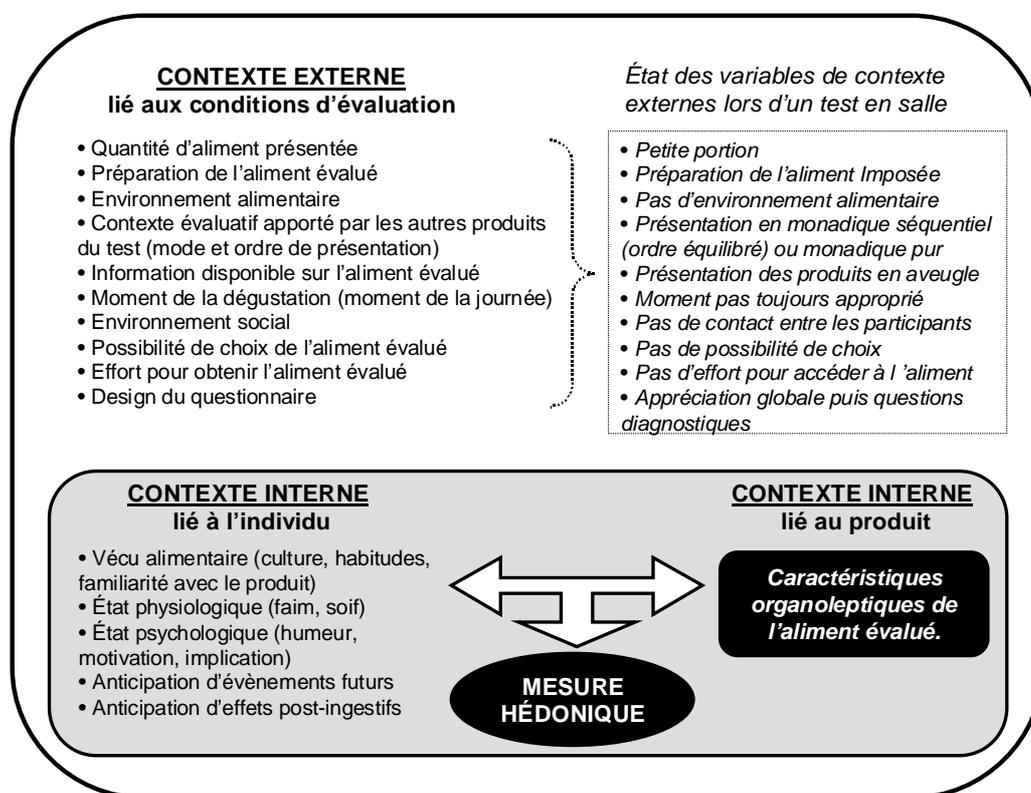


Figure 9 : Variables de contexte qui interviennent lors de l'émission d'un jugement hédonique

A ce stade de notre exposé, il nous est maintenant plus facile de comprendre pourquoi les méthodologies de test hédonique initialement développées s'attachaient à contrôler et standardiser la majorité de ces variables de contexte lors des tests consommateurs en salle ou en laboratoire afin de récupérer une mesure hédonique sensible uniquement à la variation des caractéristiques organoleptiques des produits testés. La figure 9 présente l'état des variables de contexte paramétrables lors d'un test consommateurs classique en conditions contrôlées. Cependant, la standardisation des conditions de dégustation des produits pose deux types de problèmes.

Les expérimentateurs sont loin de maîtriser la totalité des variables contextuelles qui influencent la mesure hédonique. De nombreux facteurs contextuels internes invisibles ne sont effectivement pas pris en compte dans la variabilité de la réponse mesurée. Pour Drifford *et al.* (1995), les principes opératoires des tests hédoniques qui conduisent à créer une situation isolée de tout contexte en plaçant le consommateur dans de telles conditions aseptisées induisent l'erreur de croire que les tests sont alors dénués de tout contexte. En effet, les conditions standardisées d'évaluation ne permettent pas d'éliminer le contexte mais juste de choisir un contexte plus simple (Meiselman, 2003).

Par ailleurs, les décisions stratégiques qui découlent des résultats des tests organoleptiques nécessitent des mesures hédoniques exactes. L'« exactitude » d'une méthode dans le domaine de la métrologie est définie comme « la capacité d'un instrument à produire une valeur qui est proche de la valeur exacte obtenue par une mesure indépendante qui aura été correctement calibrée » (Lawless et Heymann, 1999). Dans le domaine des sciences comportementales, ce concept d'exactitude se rapproche du critère de validité externe (ou validité prédictive) d'une méthode qui permet à l'expérimentateur de généraliser les résultats obtenus sur l'échantillon testé à la population d'intérêt. Dans le cadre plus spécifique de la mesure hédonique en test consommateurs, la validité externe concerne plus spécifiquement sa composante « écologique » qui renvoie aux conditions dans lesquelles la mesure est obtenue. Ainsi, une validité externe élevée (incluant donc une validité écologique élevée) permet de généraliser la mesure hédonique obtenue pour un échantillon de consommateurs dans des conditions de test à la population cible exposée de manière naturelle au produit. Cependant, comment imaginer que des mesures hédoniques obtenues dans de telles conditions artificielles et donc souvent inappropriées soient prédictives d'un jugement construit lors d'une vraie occasion de consommation ? Nous comprenons donc bien que la validité externe de cette mesure hédonique obtenue dans des conditions de consommation si éloignées de la réalité soit remise en question. Comme le précise Bonin *et al.* (2001), il est illusoire de croire que

l'évaluation d'un objet, qu'il soit alimentaire ou non, est strictement déterminée par ses caractéristiques intrinsèques immédiates. Ainsi, en mettant en place des tests en conditions standardisées et donc artificielles, les expérimentateurs négligent la part déterminante des facteurs contextuels dans l'évaluation des produits alimentaires (Drifford *et al.*, 1995). Köster (1981a) discute également des dangers d'une situation de test contrôlée et pointe du doigt les différences avec une situation naturelle en condamnant, entre autres, l'impossibilité pour l'individu de consommer le produit testé selon ses propres habitudes.

De manière plus globale, Schutz (1988), remettait également en question la qualité des données hédoniques classique et soulevait la nécessité d'améliorer l'efficacité des décisions prises sur la base des tests consommateurs. Il formalise (figure 10) l'influence des trois principaux composants de la mesure sensorielle (le sujet, le stimulus et le protocole de mesure) sur la qualité des données hédoniques et notamment sur leur validité externe.

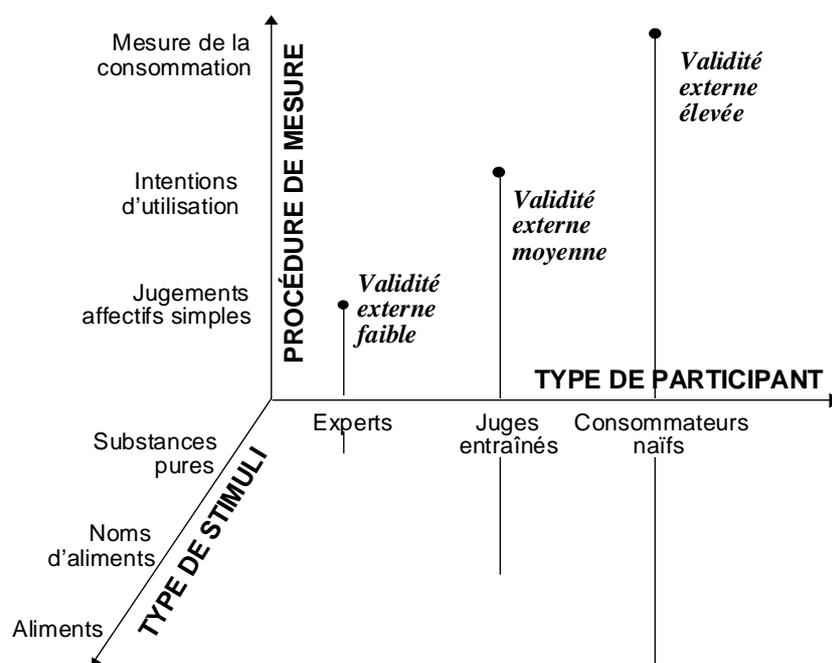


Figure 10 : Dimensions en jeu dans la validité externe des données hédoniques (d'après Schutz, 1988)

Ainsi, l'objectif d'amélioration de la validité externe des données hédoniques collectées lors des tests consommateurs supposerait l'exploration de ces trois dimensions. Si nous nous cantonnons aux exemples donnés sur la figure 10, nous pouvons tout à fait imaginer que les dimensions « type de sujet » et « type de stimuli » sont aujourd'hui bien optimisées dans la façon de mener les études consommateurs étant donné que les participants sont généralement des consommateurs « naïfs » et que les stimuli correspondent à de « vrais » aliments. Il reste

cependant sûrement encore quelques progrès à faire dans la pratique en termes de recrutement ou d'effectif de sujets. La dernière dimension concernant la procédure de mesure est en revanche beaucoup moins optimisée et consensuelle. L'augmentation du pouvoir prédictif des tests hédoniques doit donc aujourd'hui passer par une amélioration du protocole de mesure au sens large du terme (des conditions d'évaluation jusqu'au type de mesure récolté).

En 1992, après avoir fait le point sur les méthodologies employées dans la recherche sur la consommation alimentaire, Meiselman cite à nouveau le rôle de ces trois dimensions dans la qualité des données recueillies et propose de ré-orienter la recherche sur le comportement alimentaire vers une utilisation plus fréquente de « vraies » personnes consommant de « vrais » produits lors de « vraies » situations de consommation. Avec cette proposition, Meiselman remet donc en question la procédure de mesure et plus spécifiquement les conditions d'évaluation des tests en laboratoire. Il tempère son jugement en précisant que bien évidemment les études en laboratoire dans des environnements contrôlés doivent continuer d'exister. D'autres auteurs (Cardello, 1994b ; Rolls et Hetherington, 1990 ; Rolls et Shide, 1992) affirment également que les deux types de méthodologies en laboratoire et en situations naturelles sont nécessaires pour pouvoir à la fois prédire le comportement des consommateurs dans des conditions réelles (test en situation naturelle de consommation) et comprendre les raisons qui influencent ce comportement (test en situation artificielle de consommation). Par ailleurs, Garber *et al.* (2003) précisent qu'il est important de distinguer les études hédoniques théoriques, menées dans le domaine de la recherche qui peuvent supporter une faible validité externe, des études hédoniques plus appliquées à but industriel qui, en revanche, nécessitent une validation externe optimale.

**D - IMPACT DES CONDITIONS STANDARDISEES D'EVALUATION SUR LA
MESURE HEDONIQUE : CAS DES PRODUITS DANONE**

Le test en salle, tel qu'il est mis en place aujourd'hui dans des conditions de consommation standardisées et donc souvent artificielles, peut aboutir à des conclusions hédoniques déconnectées de la réalité. Cependant, l'examen de la littérature (§ C.1.2) a montré que les conditions de test standardisées n'aboutissent pas systématiquement à des résultats différents de ceux obtenus en conditions naturelles de consommation. Par ailleurs, les conclusions des études qui observent un impact des conditions de dégustation sur les résultats ne sont pas consensuelles en ce qui concerne la nature de cet impact. Il est alors difficile d'anticiper les risques que prennent les commanditaires de tests consommateurs chez Danone lorsqu'ils décident de mettre en place un test en conditions standardisées pour évaluer l'appréciation de leurs produits. Il convenait donc d'explorer l'influence des conditions artificielles de dégustation dans le cadre des méthodologies, des catégories de produit et des problématiques de test chez Danone.

Pour répondre à ces interrogations, la première partie expérimentale de cette thèse a consisté à mettre en place plusieurs tests en salle sur différentes catégories de produits afin d'identifier les spécificités des tests (produits, problématiques) pour lesquelles les conditions artificielles de consommation en salle entraînent une mauvaise validité prédictive des données collectées. Cet objectif supposait donc la possibilité de déterminer, pour chaque test, la valeur hédonique de référence. Nous avons fait le choix d'utiliser les résultats du test à domicile comme valeur hédonique de référence. Les résultats de chaque test en salle réalisé ont donc été comparés aux résultats d'un test à domicile. De fait, nous supposons que dans le cas d'obtention de résultats différents entre le test en salle et le test à domicile, le test à domicile est le garant du meilleur résultat c'est-à-dire le plus proche de la réalité. Cette hypothèse qui fait du test à domicile notre méthode de référence en termes de validation externe est particulièrement forte et est également celle d'autres auteurs (Griffin et Stauffer, 1990 ; Porcherot et Issanchou, 1998) même si elle peut être critiquée. En effet, le test à domicile n'a jamais été officiellement défini comme méthode de référence dans aucune norme. Le choix du test à domicile en tant que méthode de référence sera discuté ultérieurement dans le chapitre III qui reprend une discussion générale de ce travail.

Cette première partie expérimentale repose sur la réalisation de cinq comparaisons de tests en salle et de tests à domicile réalisés avec différentes catégories de produits Danone. Nous avons choisi de nous placer, pour chacune des études, dans une problématique de comparaison de deux ou trois produits dégustés en aveugle pour nous rapprocher des expérimentations mises en place lors de problématiques de réduction de coût, de rénovation de produit ou de comparaison par rapport à la concurrence.

Ces différentes études « *test en salle vs. test à domicile* » ont été l'occasion d'explorer différents aspects de l'influence du contexte de dégustation sur la mesure hédonique. Nous nous sommes premièrement intéressés à l'impact du mode de consommation des produits testés sur l'influence du contexte (paragraphe I). Deuxièmement, nous avons souhaité explorer l'influence des conditions de dégustation sur la mesure hédonique obtenue lors d'une problématique de réduction de sucre (paragraphe II). Enfin, une dernière expérimentation nous a permis d'explorer l'impact des caractéristiques socio-démographiques des sujets sur l'influence du contexte (paragraphe III).

Les expérimentations qui nous ont permis d'explorer ces trois voies d'investigation seront abordées séparément. Puis, un bilan présentant une discussion générale de nos résultats sur l'ensemble des expérimentations « *test en salle vs. test à domicile* » sera proposé.

I – TEST EN SALLE vs. TEST A DOMICILE : DES RESULTATS DIFFERENTS EN FONCTION DE LA CATEGORIE DE PRODUITS (ARTICLE 1)

L'illustration concernant le lien entre la catégorie de produits testée et la comparaison des résultats entre un test en salle et un test à domicile est présentée dans l'article qui suit.

Cet article présente les résultats de trois études « *test en salle vs. test à domicile* » menées en parallèle sur trois catégories différentes de produits. Ces trois études aboutissent à des conclusions différentes concernant la comparaison des données hédoniques collectées lors du test en salle ou lors du test à domicile. Ainsi, la remise en question de l'utilisation d'un test en salle ne serait pas systématique mais fonction de la catégorie du produit testé. Nous proposons alors d'interpréter la variabilité de l'influence des conditions d'évaluation sur les résultats par des hypothèses basées sur la façon dont la catégorie du produit est habituellement consommée lors d'un acte naturel de consommation.

ARTICLE 1 : Boutrolle I., Delarue J., Arranz D., Rogeaux M., Köster E. P. (2007).

Central Location Test vs. Home Use Test: contrasting results depending on product type.

Food Quality and Preference, 18, 490-499.

RESUME

Pour accéder à l'appréciation des aliments, les professionnels du marketing et de l'analyse sensorielle ont plusieurs méthodologies de test à leur disposition. Le test en salle, mis en place dans des lieux de dégustation standardisés et contrôlés est plus souvent utilisé que le test à domicile. Cependant, grâce à des conditions de consommation non imposées et donc plus réalistes, le test à domicile est supposé générer des données hédoniques plus pertinentes que les conditions artificielles de consommation du test en salle. Cet article présente la comparaison de résultats d'un test en salle et d'un test à domicile pour différentes catégories de produits. Trois expérimentations ont été mises en place avec respectivement des laits fermentés à boire, des biscuits feuilletés au fromage et des eaux gazeuses. Pour chaque test hédonique, deux variantes de produit ont été présentées aux sujets selon une procédure de présentation en monadique séquentiel. Pour chaque catégorie de produits, les échantillons obtiennent des notes d'appréciation plus élevées quand ils sont évalués à domicile qu'en salle. Nous observons également que l'influence de la méthode utilisée sur les conclusions statistiques des tests dépend de la catégorie de produits testée. En effet, alors que l'écart de note entre les deux biscuits feuilletés au fromage est influencé par la méthode, celui-ci reste identique pour les deux recettes de lait fermenté et les deux marques d'eau gazeuse. Plusieurs hypothèses basées sur la façon dont les différents produits sont habituellement consommés sont proposées afin d'expliquer la variation de l'influence des conditions d'évaluation.

Mots clés : *Test hédonique, Test en salle, Test à domicile, Contexte, Catégorie de produits, Usages, Monadique séquentiel, Monadique pur.*

ABSTRACT

Marketing professionals and sensory scientists have several hedonic testing methods at their disposal to assess product acceptability. The Central Location Test (CLT) which usually takes place in a standardized location under controlled conditions is more frequently used than the Home Use Test (HUT). Given the artificial conditions of a CLT, it is assumed that a more realistic HUT yields more relevant hedonic data in spite of the uncontrolled conditions. This study compares CLT and HUT results using different types of products. Three experiments were conducted respectively with a fermented milk beverage, salted crackers and sparkling water. For each hedonic test, two samples were presented by a monadic sequential mode. The comparison of CLT and HUT data sets shows that for each category, the products get higher scores when tested at home (HUT) than when tested under CLT conditions. Also, we found that the influence of the method used depends on the type of product to be tested. The conclusion of the hedonic test (i.e. significant difference between the two products) is affected by the method for some products (crackers) whereas the conclusion remains unchanged for other products (fermented milk beverage and sparkling water). In order to explain the varying influence of the tasting conditions, some hypotheses based on the way the products are usually eaten are discussed.

Keywords: *Hedonic test, CLT, HUT, Context, Product category, Consuming habits, Monadic sequential, Pure monadic.*

CENTRAL LOCATION TEST vs. HOME USE TEST: CONTRASTING RESULTS DEPENDING ON PRODUCT TYPE

Isabelle Boutrolle^{a,b}, Julien Delarue^b, Delphine Arranz^a, Michel Rogeaux^a, Egon Peter Köster^c

^a *Sensovaleur - Perceptions et Préférences, Danone Vitapole, RD 128, 91767 Palaiseau Cedex, France*

^b *Laboratoire de Perception Sensorielle et Sensométrie, UMR SCALE (ENSIA-CNAM-INRA), 1 avenue des Olympiades, 91744 Massy Cedex, France*

^c *Agrotechnology and Food Innovations, Wageningen University Research, Bornsesteeg 59, 6708PD Wageningen, The Netherlands*

1. INTRODUCTION

Whenever they implement innovation, improvement of the recipe or cost reduction, food companies need to assess consumers' liking for their products. Marketing professionals and sensory scientists can use various methods to measure consumer acceptance towards unbranded products. The use of hedonic testing to help in business decisions is based on the assumption that the gathered data has external validity. Nonetheless, this may be a difficult assumption, given that as Issanchou and Hossenlopp (1992) point out "the hedonic response given by a subject is only valid for the interviewed subject and the tasted product in a given context". As a result, the influence of the testing conditions on hedonic results needs to be investigated.

Today, standardized situation tests (SST) such as lab-tests or central location tests (CLT) are the most popular methods. They consist in a product evaluation by a large consumer panel after a brief exposure (one sample-subject contact) under standardised conditions of consumption. However, these declarative hedonic ratings obtained by the way of SST have not proved to be good predictors of the success or the failure of products in the market. Although the problem of unrealistic conditions of consumption in SST has been pointed out for several years (Meiselman, 1992), few alternatives have been proposed apart from the traditional home use test (HUT). This test consists of a repeated use of the product under common daily use circumstances, usually for a period of several days. However, in spite of being more realistic this method has some drawbacks such as the uncontrolled experimental conditions (e.g. the experimenters do not know who really tasted and rated the product and under what circumstances). Moreover, food manufacturers are reluctant to regularly use the HUT because of numerous logistical (i.e. short shelf life, difficulties to package blind) and budget constraints.

A number of studies compared the two approaches: standardised situation tests (SST: CLT or Laboratory tests) versus home tests (HUT or in-home testing). We found only three studies in the literature which do not show any effect of the methodology (Hersleth, Ueland, Allain, & Naes, 2005; Pound, Duizer, & McDowell, 2000; Shepherd & Griffiths, 1987). Most studies clearly show that the result of a hedonic test changes depending on the chosen methodology, but overall there is no consensus about the way in which it affects the results. Some authors report a difference in the level of liking scores with SST scores being frequently lower than HUT scores (Boutrolle, Arranz, Rogeaux, & Delarue, 2005; Daillant-Spinnler & Issanchou, 1995; Hellemann, Mela, Aaron, & Evans, 1993; Kozłowska, Jeruszka, Matuszewska, Roszkowski, Barylko-Pikielna, & Brzozowska, 2003; Murphy, Clark, & Berglund, 1958). However, in a few cases, the reverse has been observed (Daillant-Spinnler & Issanchou, 1995; Hellemann et al., 1993). In some cases, the levels of discrimination differed depending on the chosen methodology (Calvin & Sather, 1959; McDaniel & Sawyer, 1981; Miller, Nair, & Harriman, 1955). However, some contrasting results depending on the experiment prevent us from concluding that one method is more discriminating than the other. Finally, two studies report a different ranking of the products depending on whether SST or HUT was used (McEwan, 1997; Posri, MacFie, & Henson, 2001). Even when consumers evaluate their food under natural conditions, the different locations can also affect its acceptance depending on the circumstances. For example, Edwards, Meiselman, Edwards, and Leshner (2003) show that hedonic scores obtained in a test under 4 star restaurant conditions were different from those obtained in a test carried out in a student cafeteria. This pinpoints the need to investigate the impact of all environmental factors on reported hedonic results.

Natural eating situations (HUT) differ from controlled eating situations (SST) in many ways. Many of these differences could affect the reported hedonic scores. Most studies on context effects in consumer testing investigate these factors separately. Let us review the various factors that may potentially affect the results when switching from SST to HUT.

In SST, the amount of food consumed for each consumption occasion is fixed and usually small whereas it is free and usually larger during HUT. For example, Zandstra, de Graaf, van Trijp, and van Staveren (1999) observed that the optimal sucrose concentration in sweetened yoghurt is higher in an ad libitum consumption test than during a brief test. In addition, HUT allows several exposures to the sample whereas SST consists of only one brief contact. Numerous authors have shown the influence of repeated exposure on hedonic measures. Interestingly, Meiselman, de Graaf, and Leshner (2000) reported a decrease of

acceptance ratings and intake mainly due to monotony whereas Lévy (1998) observed a “mere exposure effect” (increase of acceptance of novel or complex food after repeated consumption) as described by Zajonc (1968).

The sample presentation procedure frequently differs between SST and HUT. Usually in sensory market research, mono-session monadic sequential (several products successively presented in the same session) is used for the SST whereas multi-session monadic sequential is often used for HUT (i.e. one product tasted per week). Few authors compared different presentation procedures and showed that they affect the hedonic scores (Bayton & Thomas, 1954; Lähteenmäki & Tuorila, 1994). However, no study specifically compared the mono- and the multi-session monadic sequential procedures which are those applied to the common CLT and HUT, respectively.

In HUT, subjects are free to choose the moment of consumption whereas in SST they are requested to eat during the session. As a result, the time of the day when the food item is tasted is presumably appropriate in HUT but not always in SST conditions. Birch, Billman, and Richards (1984) found that breakfast items are preferred in the morning and dinner items are preferred in the afternoon. Although these results could be expected, similar effects could also be anticipated in other situations and for other types of food. This issue is also linked to the physiological state of the subjects. Brunstrom and MacRae (1997) show that mouth-dryness influences the pleasantness of various drinks. Bell (1992) reports that hungry subjects discriminate samples less than not hungry subjects.

The HUT usually allows the consumers to choose the way they eat the products. Under the assumption of subjects' rationality we could hypothesise that in HUT they choose a personal and appropriate preparation whereas this is not always possible to achieve in SST. Matuszewska, Barylko-Pikielna, Szczecinska, and Radzanowska (1997) compared hedonic ratings of margarine samples obtained by three different tasting procedures: individual spreading on bread, prepared spread bread slices or margarine tasted alone without bread. This study shows that allowing an individual preparation of the samples led to a better discrimination. In the same way, Posri et al. (2001) demonstrated that allowing free sugar and milk doses for tea testing in a controlled environment led to more discriminating results than controlled condition. In a larger extend, the influence of the food combination was investigated by King, Weber, Meiselman, and Lv (2004). Contrary to previous authors, they showed that the discrimination was lower under the natural eating conditions: the two tested pizzas were better discriminated if they were tested alone than if tested in combination with

salad and beverages. This food environment effect must also be taken into consideration as HUT allows the combination of several food items whereas SST usually does not.

As pointed out by Stroebele and De Castro (2004), ambiance might also play a role when evaluating the products. First, it could be hypothesized that HUT conditions are more pleasant than SST conditions. Bonin, Chambres, and Bernard (2001) attempted to set up more comfortable lab test conditions but failed to show a difference in the outcomes. Other ambiance factors, especially the social interaction may play a decisive role in consumer behaviour. Thus, it may explain different hedonic results between SST and HUT since the sample consumption in SST is always individual whereas in HUT it can be social. Some studies (Berry, Beatty, & Klesges, 1985; De Castro, Brewer, Elmore, & Orozco, 1990) show that subjects eat more when eating occurred in groups of several people than when they eat alone especially when meal companions are relatives or friends (Clendenen, Herman, & Polivy, 1994).

Eventually, post-ingestional effects are probably often underestimated in the acceptability formation process. For example, HUT could allow the subject to feel post-ingestional effects before giving his/her opinion whereas CLT does not. Unfortunately, this has not received much attention in the area of food preference studies.

Several SST vs. HUT comparison studies (Calvin & Sather, 1959; Daillant-Spinnler & Issanchou, 1995; Kozłowska et al., 2003) showed that the impact of the chosen method differed depending on the kind of tested product. Accordingly, de Graaf, Cardello, Kramer, Leshner, Meiselman, and Schutz (2005) reported that the predictive ability of laboratory ratings depends on the type of food. They assumed that lab ratings are more relevant for snacks than for served dishes. We also believe that the way the product is usually eaten is a key factor in the explanation of differences between SST and HUT hedonic results. Considering the contextual factors discussed above we assume that products that are usually consumed in a large portion, as part of a family meal event or a specific situation will lead to differences between SST and HUT results. On the other hand, hedonic ratings of snacks designed for individual and convenient consumption may be less sensitive to the type of test, whether a SST or a HUT. The objective of this study was to evaluate to what extent the results of methodological comparisons between standardised situation tests and natural eating situation tests differ depending on the tested product. This paper presents three CLT vs. HUT experiments conducted in France with three different product types: fermented milk beverages, salted crackers and sparkling waters. These products differ not only by their intrinsic characteristics, but also in their ways of consumption (e.g. to drink/to eat, snack/meal

item, individual consumption/social consumption, etc.). We thus expected that these differences would affect the outcome of the hedonic tests depending on the testing conditions.

2. MATERIALS AND METHODS

Three independent CLT vs. HUT comparison experiments were carried out. The CLT and HUT protocols set up for the three experiments are those regularly used by Danone and proposed by most of the quantitative market research companies. Each experiment was carried out with separate groups of consumers (one for CLT and one for HUT).

2.1 MILK BEVERAGE EXPERIMENT

Subjects

Females between 25 and 50 year old were recruited. They were regular users of fresh dairy products and had no aversion for fermented milk beverages. Two hundred and thirty nine women (mean age 34.2 ± 7.8) took part in the CLT and two hundred and forty one women (mean age 35.2 ± 7.2) took part in the HUT.

Samples

We compared the liking for two fresh fermented milk beverages with different fat content and sugar content. One product was a regular recipe and the second one was a diet version of the same beverage. The sensory profiles carried out on the two milk beverages show that apart from the bitter taste and the metallic flavour which are more pronounced in the diet version (due to artificial sweeteners), there was no significant differences on all other attribute intensities.

2.2 SALTED CRACKERS EXPERIMENT

Subjects

Female between 18 and 65 year old were recruited. They were regular users of salted crackers and had no aversion for cheese flavour. Two hundred and forty women participated in the CLT (mean age 44 ± 12). Two hundred and forty different women participated in the HUT (mean age 44 ± 11.2).

Samples

We assessed the degree of liking for two types of salted cheese crackers: national brand vs. private label copy. The sensory profiles carried out on the two types of cheese crackers revealed many differences between these two products in terms of visual, texture and taste attributes.

2.3 SPARKLING WATER EXPERIMENT

Subjects

Females between 30 and 65 year old were recruited. They were regular users of sparkling water at ambient temperature during meals. One hundred and sixty one women took part in the CLT (mean age 45.6 ± 11.3) and one hundred and sixty women took part in the HUT (mean age 46.3 ± 10).

Samples

This experiment was carried out to assess the liking for two national brands of sparkling water: brand A vs. brand B. The sensory profiles showed that the brand B is more sparkling than the brand A and that the two brands are otherwise very similar on the other attributes.

2.4 PROTOCOL DESIGNS

Central Location Tests

Subjects were recruited in the street throughout the day (between 10.00 a.m. and 6.00 p.m.) and were invited to a testing room to participate in a 15-min tasting-session. Each subject was presented sequentially with the two samples. The sample presentation order was balanced. Subjects were instructed to eat/drink as much as they wished of each sample. Respondents evaluated the two products after having drunk some tap water. Then the subjects were asked to give a score between 1 and 10 (1=dislike very much; 10=like very much) describing their overall liking. This scale is very common for marks in the French educational system and is thus familiar to most French people. Table 1 presents some features which differ from one study to another.

Home Use Tests

The HUTs took place during two consecutive weeks. Subjects were pre-recruited by phone. The sample presentation order was balanced over the panel. Interviewers visited the participants at home three times. During the first visit, several units of the first product were given to the participants. They were free to taste the product when they wanted and together with family and/or friends but they were instructed to personally consume a minimal amount. During the second visit (six days later), the interviewers collected participants personal liking for the product (score between 1 and 10). Then they gave the consumers several units of the second product with the same instructions. After six other days, the interviewers visited the subjects again and collected the overall liking concerning the second product. Table 2 details the design of the HUT for each experiment. We distributed five conditions-of-consumption questionnaires that the subject had to fill out for five consumption occasions of the tested product, specifying the following features: when, where, with or without other food items, alone or together with other people. Unfortunately, we could not collect these usage and attitude data for the milk beverage experiment. Besides, for each experiment, we asked the consumers how much of the product they had consumed during the week.

2.5 DATA ANALYSIS

Data analysis was performed with Jump 5.0 (the SAS Institute Inc). The data set of each study was analysed separately. For all analyses, the significance level was set at 5 %.

Given the large panel sizes, we decided to analyse our data sets in two different ways for thorough analysis. On the one hand, we analysed the whole data set with two products tested per subject in a monadic sequential way (MS) i.e. related samples analysis. On the other hand, we analysed the data set as if obtained by pure monadic testing (PM). To do this, we selected only the scores of the first product tested by each subject. This second analysis allows us to suppress the sequence effect which may be different for the CLT (monadic sequential *mono*-session: 15-min between the tasting of the two products) and for the HUT (monadic sequential *multi*-session: 6 days between the tasting of the two products).

For each test, mean overall liking and standard deviations were calculated. The difference between the mean acceptability of the two products was tested by ANOVA ($score = subject + product + order$) for the MS procedure. A t-test was used for the PM procedure. Eventually, we used ANOVAs ($score = product + method + product \times method$) to determine the influence of the protocol for each tested product type.

Table 1. Summary of the testing setup for the three CLTs depending on the experiment

	MILK BEVERAGES	SALTED CRACKERS	SPARKLING WATERS
Amount of product served	1 individual blind bottle (93.7 ml)	10 units of crackers (approximately 8 g)	18 cl at ambient temperature
Container	Initial pack	White disposal plate	Transparent disposal glass
Instruction	“Drink as much as you wish but at least three mouthfuls”.	“Eat as much as you wish but at least one unit”.	“Drink as much as you wish but at least half of the glass”
Administration	Interview	Self administered	Self administered

Table 2. Summary of the testing setup for the three HUTs depending on the experiment

	MILK BEVERAGES	SALTED CRACKERS	SPARKLING WATERS
Amount of product distributed	6 small blind bottles (6 × 93.7 ml)	2 large blank plastic pack (2 × 100 g)	5 bottles (5 × 500 ml)
Instruction	“Drink as much as you wish but at least 5 bottles personally”	“Eat as much as you wish but at least 1 pack personally”	“Drink as much as you wish but at least 4 bottles personally”
Conditions-of-consumption questionnaire to complete for five consumption occasions.	None	Testing conditions questionnaires: when, where, with or without other food items, alone or together with other people.	Testing conditions questionnaires: when, where, with or without other food items, alone or together with other people.
Administration	Interview at home 6 days after product delivery	Self-administration when the subject begins to eat the biscuits of the second packet.	Self-administration when the subject begins to drink the fourth bottle.

3. RESULTS

3.1. FERMENTED MILK BEVERAGE EXPERIMENT

Table 3 summarises the mean scores analyzed in a MS way for each fermented milk beverage for CLT and for HUT. Here, CLT and HUT lead to the same conclusion of a significant *product* effect. The mean acceptability is significantly higher for the standard beverage than for the diet beverage. However, it should be noted that the ANOVA shows a significant *method* effect [$F_{(1, 956)} = 7.574$; $p < 0.05$]. Indeed, the scores obtained in HUT are significantly higher than those obtained in CLT conditions. At the same time, the *product* × *method* interaction is not significant [$F_{(1, 956)} = 0.237$; $p > 0.05$] which indicates that even if the CLT mean scores are different from the HUT mean scores, the product ranking order and the differences between the mean scores of the two products are similar for the two methods. Table 4 summarises the mean scores analyzed in a PM way for each fermented milk beverage for CLT and HUT. Here, we observe that the consumers like the two versions of the beverage equally well in both the CLT and HUT. The ANOVA shows a non significant *method* effect [$F_{(1, 476)} = 0.081$; $p > 0.05$] and a non significant *product* × *method* interaction [$F_{(1, 476)} = 0.001$; $p > 0.05$].

To summarise, the MS and PM analyses lead to different conclusions (significant difference or not). Besides, the method affects the degree of hedonic ratings only in the case of the MS design. However, in both cases there is no impact of the use of CLT or HUT on the difference in liking between the two tested milk beverages (the *product* × *method* interaction is not significant). Thus given this result, we can hypothesise that for the fermented milk beverage experiment, the CLT evaluation conditions (only one exposure to each product under standardised conditions) are sufficient to collect consumers' liking that is similar to those obtained after one week of home consumption.

Table 3. Mean scores analyzed in a monadic sequential way for CLT and HUT.

		MILK BEVERAGES		SALTED CRACKERS		SPARKLING WATERS	
		Standard	Diet	National brand	Private label	Brand A	Brand B
CLT		239		240		161	
	Panel size						
	Mean score (SD)	6.96 (1.90)	6.55 (2.13)	7.14 (1.93)	7.02 (1.97)	6.74 (1.77)	6.20 (1.98)
	Subject effect	$F_{(238,237)}=2.43; p<0.001$ **		$F_{(239,238)}=1.72; p<0.001$ **		$F_{(160,159)}=1.21; p=0.11$ NS	
	Product effect	$F_{(1,237)}=8.53; p<0.01$ **		$F_{(1,238)}=0.74; p=0.39$ NS		$F_{(1,159)}=7.38; p<0.01$ **	
	Order effect	$F_{(1,237)}=38.63; p<0.001$ **		$F_{(1,238)}=1.45; p=0.23$ NS		$F_{(1,159)}=0.96; p=0.33$ NS	
HUT		241		240		160	
	Panel size						
	Mean score (SD)	7.24 (1.79)	6.95 (1.78)	7.91 (1.37)	7.09 (1.89)	7.54 (1.62)	7.14 (1.82)
	Subject effect	$F_{(240,239)}=2.09; p<0.001$ **		$F_{(239,238)}=1.28; p<0.05$ *		$F_{(159,158)}=1.11; p=0.26$ NS	
	Product effect	$F_{(1,239)}=4.81; p<0.05$ *		$F_{(1,238)}=35.15; p<0.001$ **		$F_{(1,158)}=4.61; p<0.05$ *	
	Order effect	$F_{(1,239)}=0.46; p=0.50$ NS		$F_{(1,238)}=15.62; p<0.001$ **		$F_{(1,158)}=2.02; p=0.16$ NS	

Anova results (score = subject + product + order): NS: not significant; *: p-value < 5 %; **: p-value < 1 %.

Table 4. Mean scores analyzed in a pure monadic way (scores of the first tasted sample only) for CLT and HUT.

		MILK BEVERAGES		SALTED CRACKERS		SPARKLING WATERS	
		Standard	Diet	National brand	Private label	Brand A	Brand B
CLT		120		120		80	
	Panel size						
	Mean score	7.22	7.15	7.51	7.42	6.80	6.33
	SD	1.80	1.84	1.75	1.63	1.84	1.98
HUT		121		120		80	
	Panel size						
	Mean score	7.17	7.10	8.07	7.47	7.63	7.33
	SD	1.94	1.63	1.22	1.60	1.65	1.74

t-test results: NS: not significant; *: p-value < 5 %; **: p-value < 1 %.

3.2. SALTED CRACKERS EXPERIMENT

As can be seen in Table 3, the conclusion of the test depends on the used method. The HUT data set gives rise to a significant difference between the two mean hedonic scores: the mean acceptability of the national brand is significantly higher than that of the private label. Contrary to the HUT scores, the CLT scores show a non significant difference between the national brand and the private label.

Just as in the milk beverage experiment, the mean hedonic score obtained for the two types of salted crackers in the HUT is higher than that obtained in the CLT. This phenomenon is demonstrated by the significant *method* effects found in the ANOVA [$F_{(1, 956)} = 12.869$; $p < 0.05$]. As could be expected on the basis of the different level of discrimination between the CLT and the HUT, the *product × method* interaction is significant [$F_{(1, 956)} = 29.751$; $p < 0.05$].

Table 4 shows the results obtained for the PM analysis. The comparison of CLT and HUT results lead to the same observation as the MS analysis: the conclusion of the test depends on the used method. In HUT, overall liking of the national brand is significantly higher than that of the private label whereas in CLT the two brands are equally liked. Again, the results of the ANOVA show a significant *method* effect [$F_{(1, 476)} = 11.408$; $p < 0.05$] and a marginally significant *product × method* interaction at a 10 % level [$F_{(1, 476)} = 3.287$; $p < 0.1$]. The salted crackers experiment leads to contrasted results with the milk beverage experiment. Indeed, the conclusions obtained by the way of the MS analysis or the PM analysis are identical in the salted crackers experiment and this is not the case in the milk beverage experiment. Besides, the salted crackers experiment shows an impact of the used method whereas the milk beverage experiment does not. A sensory profile carried out on the crackers before the hedonic tests (data not presented here) showed that the two tested products were very different for several sensory attributes. Consequently, we could hypothesise that the non-significant difference of acceptability obtained in CLT does not come from the fact that the consumers had difficulties to differentiate the products. Thus, for the salted crackers tested in this experiment, the evaluation conditions influence not only the level of the hedonic scores but also the conclusion inferred from the tests, which is more worrying. The HUT results reveal a significant difference of acceptability between the two types of tested crackers that the CLT results do not suggest. Given these results, we can conclude that for the two types of salted crackers, the not significant difference in liking obtained under CLT evaluation conditions does not represent the possibly "true" difference in liking that probably exists.

3.3. SPARKLING WATERS EXPERIMENT

Table 3 presents the mean scores analyzed in a MS way for each sparkling water for CLT and for HUT. We can see the same conclusion whatever the used method: the mean acceptability is significantly higher for brand A than for brand B. The scores obtained in HUT are significantly higher than those obtained in CLT conditions. This is illustrated by the significant *method* effect [$F_{(1,638)} = 37.13$; $p < 0.05$] found in the ANOVA. However, the *product* \times *method* interaction is not significant [$F_{(1,638)} = 0.22$; $p > 0.05$]. The product ranking order and the differences between the mean scores of the two waters are similar for CLT and HUT.

Table 4 presents the results of the PM analysis. Contrary to the MS data, the pure monadic procedures both for CLT and for HUT are not able to show a difference of liking. Thus, the consumers like the two waters equally well in both the CLT and HUT. The ANOVA shows a significant *method* effect [$F_{(1,317)} = 20.28$; $p < 0.05$] and a non significant *product* \times *method* interaction [$F_{(1,317)} = 0.17$; $p > 0.05$].

At first glance, these results look like those of the milk beverages. The MS data sets lead to the same conclusion for CLT and HUT. To conclude, even if the MS and PM analyses lead to different results, the impact of the tasting conditions on this kind of products was not observed in this experiment. Here we did not observe major differences between CLT and HUT results. Therefore, it could be assumed that for this kind of products, the CLT provides “true” hedonic results. However, this point is discussed in the following.

4. DISCUSSION

4.1 PURE MONADIC vs. MONADIC SEQUENTIAL COMPARISONS

The results of the present study show that the mean scores obtained according to the PM analyses are often slightly higher than those obtained in the MS procedure, regardless of the experiment and of the testing conditions. This may logically result from the well-known over-scoring effect for the first tasted product (Stone & Sidel, 1993), given that in the PM analysis we only take the score obtained by the first evaluated product into account. In order to quantify this phenomenon, we can examine the order effect in each ANOVA carried out on the six data sets (Table 3). It can be seen that only two out of six trials show a significant order effect: the CLT of the fermented milk beverages and the HUT of the salted crackers. Yet this first-sample effect should not be neglected. Some authors consider it to be an artefact

resulting from the positional bias and use a “warm up” sample (Forde & Delahunty, 2004; Hersleth, Mevik, Naes, & Guinard, 2003; Jaeger, Andani, Wakeling, & MacFie, 1998). On the other hand, it could be argued that in the MS procedure, the hedonic judgement of the second sample is biased as it implies a comparison with the first evaluated sample. In this respect, the PM procedure is expected to lead to a more natural judgment because it allows the subject to score the tasted product relatively to his/her internal reference such as his/her usual or favourite product.

Besides, our results show that in the case of the milk beverages and of the sparkling waters, the conclusions inferred from the MS analysis are different from those of the PM analysis whatever the testing conditions. Indeed, the MS analysis leads to a significant product effect whereas the PM analysis leads to a non significant difference between the mean hedonic scores. First it should be noted that the PM analysis is based on a data set which is mechanically two times smaller than the initial MS data set. Therefore, the discrepancy between the two analyses carried out with two different sample sizes was not unexpected. This is a purely statistical effect which directly results from a decrease in the power of any difference test when the sample size decreases. The double sample size of the sequential procedure provides a real advantage in order to find a significant effect when the difference between the mean scores is small. In other words and to the exclusion of any sensory and psychological considerations, finding a difference of acceptability is easier with MS than with PM with the same overall number of subjects. This may have practical implications when the cost of the test, which is highly dependent on the number of interviewed consumers, is a limiting constraint.

Nevertheless, the difference between MS and PM results could also be ascribed to the sensory evaluation paradigm. Indeed, given the close sensory similarity between the two tested samples, consumers may need to taste the two products one after the other in order to possibly detect a difference. This could explain the fact that the monadic sequential and the pure monadic analyses lead to the same conclusions in the case of the salted crackers experiment for which the two products have very different sensory profiles (data not shown), whereas the test conclusions differ in the case of the milk beverage and the sparkling water experiments where the sensory profiles of the two compared products were more similar.

More generally, relative hedonic scores (comparative procedures) are sometimes presented as being more sensitive than absolute hedonic scores (pure monadic) (Bayton & Thomas, 1954; McBride, 1982). This is confirmed by subsequent data analysis presented in table 5, only taking into account the hedonic scores of the “second tasted sample”. In the case

of the milk beverage experiment for example, the Student's tests carried out on the "second sample" sub-dataset show a significant difference of liking between the two beverages, both for the CLT ($t = 10.0$; $p < 0.05$) and for the HUT ($t = 9.52$; $p < 0.05$), whereas the "first sample" (MP) sub-dataset does not result in a significant difference either for the CLT ($t = 1.17$; $p > 0.05$) or for the HUT ($t = 1.41$; $p > 0.05$). Considering these results, the use of a warm up sample in hedonic testing would be an interesting option when a higher sensitivity or higher discrimination is sought, which is by the way not always granted.

Table 5. Mean scores analysis only taking into account the scores of the "second tasted sample" for CLT and HUT.

	MILK BEVERAGES		SALTED CRACKERS		SPARKLING WATERS	
	Standard	Diet	National brand	Private label	Brand A	Brand B
CLT						
Panel size	119	120	120	120	81	80
Mean score	6.70	** 5.96	6.78	NS 6.62	6.68	* 6.06
SD	1.95	2.24	2.04	2.20	1.72	1.98
HUT						
Panel size	121	120	120	120	80	80
Mean score	7.30	* 6.80	7.75	** 6.70	7.46	NS $p=0.07$ 6.95
SD	1.62	1.92	1.50	2.07	1.66	1.89

t-test results: NS: not significant; *: p-value < 5 %; **: p-value < 1 %.

4.2. CLT vs. HUT COMPARISONS

It is striking that, in all our experiments, the products get higher scores when tested at home (HUT) than when tested in CLT conditions. This is confirmed by the significant *method* effects in the respective ANOVAs with the exception of the milk beverages in the PM analysis. Previous studies also showed that standardised situation tests underestimate product acceptance as compared to tests carried out in natural settings (Hersleth et al., 2003; King et al., 2004; Kozłowska et al., 2003; Meiselman, Hirsch, & Popper, 1988; Meiselman, Johnson, Reeve, & Crouch, 2000; Murphy et al., 1958). Several differences between CLT and HUT could account for this phenomenon.

First, care should be taken to the fact that in our experiments, the *method* effect (higher scores in HUT) is confounded with the *panel* effect (consumer samples are different). Indeed, we cannot completely exclude the possibility that the *method* effect found in our three experiments could be due to the fact that the HUT panels and CLT panels are different.

However considering the relatively large size of our consumer samples, this *panel* effect could reasonably be neglected.

The most widely spread assumption is that the more natural evaluation conditions at home favourably influence the overall well-being of the participants and therefore their overall liking. Although this has never been proved experimentally, it would be interesting to check this assumption by assessing the mood states of the interviewed subjects in the two different conditions, using for example the DES (Differential Emotions Scale) developed by Izard, Dougherty, Bloxom, and Kotsch (1974).

More likely, the consumption conditions in HUT are expected to be more appropriate than CLT conditions. During the HUT, the subjects choose when they want to consume the products. Thus, the time of the day is assumed to be appropriate to the consumption of the products (breakfast for milk beverages, evenings for salted crackers and meal for sparkling waters) which is not systematically the case during the CLT. Moreover, the subjects are supposed to eat/drink when they feel hungry or thirsty. Thus, these two features of the HUT might naturally improve the subjects' overall satisfaction. The prolonged contact with the products in HUT could also explain the increase of liking scores. Indeed, Zajonc (1968) showed an improvement of the product liking scores due to familiarisation with the product when the subject – sample contact duration increases (mere exposure effect). However, we do not know whether this mere exposure effect could occur for the three kinds of products that we tested.

It is also very likely that the subjects were not involved in the evaluation task in the same way depending on the test conditions, which could possibly result in a cognitive bias. We indeed believe that the CLT standardised conditions reinforced the sensation of a formal experiment and that the subjects were placed in an analytical mindset and were more likely to set themselves the task of “analysing” the samples. Thus, these formal conditions might have led them to be more critical and demanding towards the tested products. Such a bias has already been found for the evaluation of attribute intensities by consumers (Pound et al., 2000). Besides, in the case of the HUT, it is usually impossible to check if the subject has strictly followed the instructions. From our experience, we suspect that some subjects are less involved in the test or simply forget to taste the required amount of product. If so, we cannot completely exclude that some subjects evaluated the product at the very end of the test period, just before the interview or the recovering of the questionnaires. As such subjects might feel guilty; they might be more lenient towards the products and give higher scores.

Most interestingly, the main information revealed by these experiments is that, apart from the underscoring of the products in CLT, the influence of the used method depends on the experiment. Before suggesting explanations for these contrasting results, it should be noted that two experiments were carried out in face-to-face interviews (crackers and sparkling water) whereas the third was self-administered (milk beverages). Naturally, this might have affected the way the subjects scored the products. Nevertheless, this does not affect the CLT vs HUT comparisons since for one product type, the protocol remained the same. In addition to this, the results of the CLT vs HUT comparisons differed between crackers and sparkling water experiments which were both carried out in face-to-face interviews. This confirms the fact that the way the interviews were carried out does not explain the influence of the method used in each experiment. We can thus assert with some confidence that the influence of the used method depends on the type of product to be tested. The conclusion of the test (i.e. significant difference between the two products) is indeed affected by the method for some products (crackers) whereas the results remain unchanged for other products (milk beverages, sparkling water). Naturally, the three types of tested products may imply different ways of consumption. As previously mentioned, numerous studies revealed the role of these consumption conditions: amount consumed, time of the day, combination with other food items and social context. In our case, it is thus likely that the way the products are usually eaten could explain the varying influence of the tasting conditions (HUT and CLT). For the crackers and the sparkling waters, the way the samples were consumed during the HUT can be analysed thanks to the conditions-of-consumption questionnaires. In the case of the fermented milk beverages, a Usage & Attitude survey previously carried out by the marketing research services of Danone gives us some insights into the way the products are usually consumed.

The fermented milk beverage survey showed that 64% of French people consume this product in the morning during breakfast. This product is otherwise consumed as a snack throughout the day. Considering that the average time duration for breakfast is only of about fifteen minutes and that French eat alone for breakfast in fifty per cent of cases, we can state that the conditions of consumption for this kind of fermented milk beverages are not very convivial and social. We thus assume that these consumption conditions are not very different from the CLT conditions, except for the time of the day. Even if the total amount of product consumed in HUT and CLT is not the same (5 small bottles vs. 1 small bottle), the amount of product consumed at each consumption occasion is the same in both conditions since the product is packaged in small individual bottles. This implies that consumers drink the same

portion (93.7 ml) every time. It is thus not surprising that the two protocols led to the same hedonic results.

In the case of the salted crackers HUT, the conditions-of-consumption questionnaires indicate that about 70% of the consumption occasions occurred early in the evening in a social context with family or friends, typically before diner for “apéritif”. Naturally, this situation implies that these products were eaten in combination with beverages (56% with alcohol, 32% with fruit juice or soda). Thus, the conditions of consumption for the salted crackers during the HUT were rather social and convivial and took place in a relaxing atmosphere, clearly a different situation than during the CLT. Besides, the total amount of eaten crackers in HUT was rather large for both samples (more than 80 % of the subjects ate at least 1 pack of crackers, i.e. 100g). On the contrary, we observed that during the CLT, the subjects ate only few units of crackers (and rarely more than 1 unit). Here it is thus not surprising that CLT and HUT led to different results.

In the case of the sparkling water, we had hypothesised that these products were usually consumed as part of a meal event. This is indeed the way they are marketed. We thus expected to get different results from HUT and CLT, as it was the case for the crackers. However, the two conditions led to the same conclusions. We thus explored the questionnaires detailing the conditions of consumption during the HUT. Contrary to our assumption, we found out that about 40% of the consumption occasions occurred between meals. Thus, only 60% of HUT consumption occasions occurred during a meal event. Therefore, this behaviour might have somewhat balanced the results.

The results of the crackers experiment and of the milk beverage experiment lead us to conclude that the usual context of consumption has a leading role in the hedonic evaluation of food products. This may have important implications in the choice of a test methodology, i.e. HUT or CLT. We could advise the use of CLT, which is supposed to be more convenient, when the products to be tested are not strongly related to a specific consumption context. Conversely, HUT should be used whenever the product consumption is expected to be very contextual. Accordingly, it would be interesting to test the influence of context variables in controlled testing conditions or situational CLT as King et al. (2004) did. Systematic studies in this direction would be needed to better identify the most influencing context variables that should not be neglected when implementing a CLT. In the case of the sparkling water, whose conditions of consumption were somewhat balanced between meal-context and out-of-meal context, the test method did not affect the results (apart from the usual underscoring effect in CLT). This underlines the need to work on the concept of appropriateness for the hedonic

testing of food products, especially those which are expected to be multi-situational (Cardello & Schutz, 1996). Furthermore, our experience in CLT implementation highlighted a lack of involvement of the participants in the test. Numerous CLT features, declarative data especially, have a great chance to bring about casual responses and results quite far from the actual consumption or purchasing behaviour. Given this consideration, the use of incentive compatible hedonic measurement methods developed in experimental economics is potentially an interesting improvement for SST methodologies.

REFERENCES

- Bayton, J. A., & Thomas, C. M. (1954). Comparative single stimulus methods determining taste preferences. *Journal of Applied Psychology*, 38, 443-445.
- Bell, J. A. (1992). The relationship of hedonics to behavior. Limitations and future of acceptance testing. *Activities Report of the RED Associates*, 44, 138-150.
- Berry, S. L., Beatty, W. W., & Klesges, R. C. (1985). Sensory and social influences on ice cream consumption by males and females in a laboratory setting. *Appetite*, 6, 41-45.
- Birch, L. L., Billman, J. B., & Richards, S. S. (1984). Time of day influences food acceptability. *Appetite*, 5, 109-116.
- Bonin, D., Chambres, P., & Bernard, P. (2001). Influence du contexte sur la perception des produits alimentaires. In: I. Urdapilleta, C. Ton Nu, & C. Saint Denis (Eds.), *Traité d'évaluation sensorielle* (pp. 223-232). Paris: Dunos.
- Boutrolle, I., Arranz, D., Rogeaux, M., & Delarue, J. (2005). Comparing central location test and home use test results: Application of a new criterion. *Food Quality and Preference*, 16, 704-713.
- Brunstrom, J. M., & MacRae, A. W. (1997). Mouth state: A nuisance variable in preference tests? *Food Quality and Preference*, 8, 349-352.
- Calvin, L. D., & Sather, L. A. (1959). A comparison of student preference panels with a household consumer panel. *Food Technology*, 13, 469-472.
- Cardello, A. V., & Schutz, H. G. (1996). Food appropriateness measures as an adjunct to consumer preference/acceptability evaluation. *Food Quality and Preference*, 7, 239-249.
- Clendenen, V. I., Herman, C. P., & Polivy, J. (1994). Social Facilitation of Eating Among Friends and Strangers. *Appetite*, 23, 1-13.
- Dailliant-Spinnler, B., & Issanchou, S. (1995). Influence of label and location of testing on acceptability of cream cheese varying in fat content. *Appetite*, 24, 101-105.
- De Castro, J. M., Brewer, E. M., Elmore, D. K., & Orozco, S. (1990). Social facilitation of the spontaneous meal size of humans occurs regardless of time, place, alcohol or snacks. *Appetite*, 15, 89-101.
- De Graaf, C., Cardello, A. V., Kramer, F., Leshner, L. L., Meiselman, H. L., & Schutz, H. G. (2005). A comparison between liking ratings obtained under laboratory and field conditions: the role of choice. *Appetite*, 44, 15-22.
- Edwards, J. S. A., Meiselman, H. L., Edwards, A., & Leshner, L. (2003). The influence of eating location on the acceptability of identically prepared foods. *Food Quality and Preference*, 14, 647-652.
- Forde, C. G., & Delahunty, C. M. (2004). Understanding the role cross-modal sensory interactions play in food acceptability in younger and older consumers. *Food Quality and Preference*, 15, 715-727.
- Helleman, U. A., Mela, D. J., Aaron, J. I., & Evans, R. E. (1993). Role of fat in meal acceptance. *Food Quality and Preference*, 4, 90.
- Hersleth, M., Mevik, B.H., Naes, T., & Guinard, J.X. (2003). Effect of contextual factors on liking for wine - use of robust design methodology. *Food Quality and Preference*, 14, 615-622.
- Hersleth, M., Ueland, O., Allain, H., & Naes, T. (2005). Consumer acceptance of cheese, influence of different testing conditions. *Food Quality and Preference*, 16, 103-110.
- Issanchou, S., & Hossenlopp, J. (1992). Les mesures hédoniques: méthodes, portées et limites. In: I. Giachetti (Ed.), *Plaisirs et Préférences Alimentaires* (pp. 49-75). Paris: Polytechnica.

- Izard, C. E., Dougherty, F. E., Bloxom, B. M., & Kotsch, W. E. (1974). *The differential emotions scale: A method of measuring the subjective experience of discrete emotions*. Department of Psychology. Nashville, TN: Vanderbilt University.
- Jaeger, S. R., Andani, Z., Wakeling, I. N., & MacFie, H. J. H. (1998). Consumer preferences for fresh and aged apples: a cross-cultural comparison. *Food Quality and Preference, 9*, 355-366.
- King, S. C., Weber, A. J., Meiselman, H. L., & Lv, N. (2004). The effect of meal situation, social interaction, physical environment and choice on food acceptability. *Food Quality and Preference, 15*, 645-653.
- Kozłowska, K., Jeruszka, M., Matuszewska, I., Roszkowski, W., Barylko-Pikielna, N., & Brzozowska, A. (2003). Hedonic tests in different locations as predictors of apple juice consumption at home in elderly and young subjects. *Food Quality and Preference, 14*, 653-661.
- Lähteenmäki, L., & Tuorila, H. (1994). Liking for ice cream measured with three procedures: side-by-side, after consumption and single samples. *Journal of Sensory Studies, 9*, 455-465.
- Lévy, C. (1998). *Elaboration de méthodologies adaptées à l'étude de la dynamique des préférences*. Doctoral thesis, Université de Bourgogne, Dijon, France.
- Matuszewska, I., Barylko-Pikielna, N., Szczecinska, A., & Radzanowska, J. (1997). Comparison of three procedures for consumer assessment of fat spreads: short report. *Polish Journal of Food and Nutrition Sciences, 6*, 139-142.
- McBride, R. L. (1982). Range bias in sensory evaluation. *Journal of Food Technology, 17*, 405-410.
- McDaniel, M. R., & Sawyer, F. M. (1981). Preference testing of whiskey sour formulations: magnitude estimation versus the 9-point hedonic. *Journal of Food Science, 46*, 182-185.
- McEwan, J. A. (1997). A comparative study of three product acceptability trials. *Food Quality and Preference, 8*, 183-190.
- Meiselman, H. L. (1992). Methodology and theory in human eating research. *Appetite, 19*, 49-55.
- Meiselman, H. L., deGraaf, C., & Leshner, L. L. (2000). The effects of variety and monotony on food acceptance and intake at a midday meal. *Physiology & Behavior, 70*, 119-125.
- Meiselman, H. L., Hirsch, E. S., & Popper, R. D. (1988). Sensory, hedonic and situational factors in food acceptance consumption. In: D. M. H. Thomson (Ed.), *Food acceptability* (pp. 77-87). London: Elsevier.
- Meiselman, H. L., Johnson, J. L., Reeve, W., & Crouch, J. E. (2000). Demonstrations of the influence of the eating environment on food acceptance. *Appetite, 35*, 231-237.
- Miller, P. G., Nair, J. H., & Harriman, A. J. (1955). A household and a laboratory type of panel for testing consumer preference. *Food Technology, 9*, 445-449.
- Murphy, E. F., Clark, B. S., & Berglund, R. M. (1958). A consumer survey versus panel testing for acceptance evaluation of Maine sardines. *Food Technology, 12*, 222-226.
- Posri, W., MacFie, H., & Henson, S. (2001). Improving the predictability of consumer liking for central location test in tea. In: *4th R.M. Pangborn memorial symposium: A sense odyssey*. Dijon, France, 22-26 July.
- Pound, C., Duizer, L., & McDowell, K. (2000). Improved consumer product development. Part one: Is a laboratory necessary to assess consumer opinion? *British Food Journal, 102*, 810-820.
- Shepherd, R., & Griffiths, N. M. (1987). Preferences for eggs produced under different systems assessed by consumer and laboratory panels. *Lebensmittel Wissenschaft und Technologie, 20*, 128-132.
- Stone, H., & Sidel, J. L. (1993). *Sensory Evaluation Practices (2d ed.)* San Diego, California: Academic Press Inc.
- Stroebele, N., & De Castro, J. M. (2004). Effect of ambience on food intake and food choice. *Nutrition, 20*, 821-838.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1-27.
- Zandstra, E. H., de Graaf, C., van Trijp, H. C. M., & van Staveren, W. A. (1999). Laboratory hedonic ratings as predictors of consumption. *Food Quality and Preference, 10*, 411-418.

II – TEST EN SALLE vs. TEST A DOMICILE : CAS D'UNE PROBLEMATIQUE DE REDUCTION DE SUCRE

Dans le contexte actuel de l'obésité croissante, de nouvelles réglementations sont imposées aux entreprises agroalimentaires afin de diminuer la teneur en sucre des produits qu'elles commercialisent. Ainsi, de nombreux projets de rénovation ayant comme objectif de développer des produits moins sucrés voient le jour quotidiennement chez Danone afin d'évaluer les conséquences de la diminution de la teneur en sucre des produits sur leur appréciation. La littérature sur l'influence des variables de contexte a mis en évidence un impact de la quantité consommée et du nombre de contacts entre les sujets et les produits sur la détermination de la teneur en sucre ou en sel idéale pour les consommateurs. Dans un tel contexte il nous a semblé important d'évaluer l'efficacité des tests en salle proposant la consommation ponctuelle d'une faible quantité de produit pour révéler l'appréciation de produits à teneurs en sucre différentes. Nous avons donc testé la validité externe d'un test en salle traditionnel pour évaluer l'appréciation de trois recettes de boissons à base de jus de fruit et de lait ayant différentes teneurs en sucre. Cette validité a été testée par comparaison à un test à domicile proposant la consommation répétée d'une quantité importante de boisson.

II.1 - MATERIELS ET METHODES

II.1.1 - Les produits

L'appréciation de trois variantes de recettes d'un jus de fruit saveur pêche abricot mélangé avec du lait a été mesurée :

- la recette standard (A),
- une recette avec 25% de sucre ajouté en moins (B),
- une recette avec 100% de sucre ajouté en moins (C).

II.1.2 - Les sujets

Au total, 1002 personnes ont été interrogées. Quatre cent quatre-vingt-six personnes ont été interrogées dans les conditions d'un test en salle et 516 personnes ont été interrogées selon les conditions d'un test à domicile. Les évaluations ayant été réalisées en monadique pur, chaque sujet n'a dégusté qu'un seul produit. La répartition des sujets selon leur âge et leur sexe est la suivante :

- 50% d'hommes et 50% de femmes,
- 50% de personnes entre 20 et 39 ans et 50% de personnes entre 40 et 69 ans.

Tous les sujets interviewés sont des consommateurs de la marque d'intérêt (mais pas forcément du parfum testé) à la fréquence minimum d'une occasion de consommation tous les deux mois.

II.1.3 - Les protocoles

Le test en salle

Les sujets sont recrutés dans les rues de six villes différentes tout au long de la journée et sont invités à participer à une séance de dégustation. Chaque sujet ne déguste qu'un seul produit présenté frais dans un verre en plastique. Le sujet boit la quantité de boisson qu'il souhaite puis doit donner une note d'appréciation globale au produit en utilisant une échelle allant de 1 à 10. La séance se poursuit ensuite par un ensemble de questions¹ basées sur des caractéristiques spécifiques du produit avec, entre autres, une question de notation par rapport à l'idéal (JAR) de l'intensité sucrée des boissons.

Le test à domicile

Les sujets sont pré-recrutés dans les six mêmes villes que pour le test en salle. Ici encore, chaque sujet ne déguste qu'un seul des produits. La première visite de l'enquêteur au domicile du sujet consiste à distribuer deux bouteilles d'un litre de la boisson testée. Le sujet a alors pour instruction de conserver les boissons au réfrigérateur et de consommer au moins personnellement une des deux bouteilles comme il le souhaite et à la fréquence qu'il le souhaite avant de donner son appréciation globale (note sur 10) sur un questionnaire auto-administré². L'interviewé doit également remplir un carnet de consommation³ pour chaque occasion de consommation dans lequel il détaille les conditions dans lesquelles il a consommé le produit (quand, où, avec qui, avec quoi). Au bout de cinq ou six jours, l'enquêteur revient au domicile de l'interviewé pour récupérer le questionnaire d'appréciation globale et le carnet de consommation puis il lui pose les questions diagnostiques relatives aux caractéristiques spécifiques du produit⁴.

¹ voir Annexe IV-a

² voir Annexe IV-b

³ voir Annexe IV-c

⁴ voir Annexe IV-d

II.1.4 - Analyse des données

Pour chacune des méthodes, les notes moyennes d'appréciation globale ainsi que les écart-types pour chaque recette sont calculés. La différence d'appréciation entre les trois produits est évaluée par l'intermédiaire d'une ANOVA à un facteur (*note = produit*). Un test de Newman-Keuls permet ensuite de déterminer les recettes dont les appréciations sont significativement différentes. Une ANOVA (*note = produit + méthode + produit × méthode*) est ensuite réalisée de manière à déterminer l'influence des conditions d'évaluation sur l'appréciation globale collectée.

L'influence de la méthode sur les réponses à la question JAR concernant l'intensité sucrée est évaluée pour chacune des recettes par un test du χ^2 qui compare les distributions des réponses selon la méthode.

II.2 - RESULTATS

II.2.1 - Influence du contexte sur l'appréciation globale des produits

Le tableau 2 présente les notes moyennes obtenues pour chaque produit selon la méthodologie employée ainsi que les résultats de l'ANOVA pour chaque test.

Tableau 2 : Notes moyennes obtenues pour chaque recette selon la méthode et résultats de l'ANOVA. (*note = produit*) pour chaque test.

	Test en salle			Test à domicile		
	Recette A	Recette B	Recette C	Recette A	Recette B	Recette C
Effectif	160	160	160	176	170	170
Note moyenne	7,70 ^a	6,81 ^b	6,28 ^c	8,07 ^a	7,80 ^a	7,37 ^b
Ecart-type	1,48	1,60	1,87	1,37	1,61	1,68
Effet produit	F _(2,483) = 30,59 (p < 0,001)			F _(2,513) = 8,84 (p < 0,001)		

Le classement des recettes selon leur appréciation illustre le goût prononcé des sujets pour les produits sucrés. En effet, quelle que soit la méthode, la recette avec la plus forte teneur en sucre est la recette la plus appréciée suivie par la recette avec 25% de sucre ajouté en moins puis par la recette sans sucre ajouté. L'observation de ce tableau nous permet d'aboutir à des premières conclusions en ce qui concerne la comparaison du test en salle et du test à domicile. Premièrement, nous observons, une fois de plus, que les notes d'appréciation globale collectées à domicile sont plus élevées que les notes collectées en salle. Ceci est

confirmé par l'effet *méthode* significatif [$F_{(1,996)} = 64,36$; $p < 0,001$] donné par l'ANOVA. Par ailleurs, nous constatons également que les conclusions concernant la comparaison de l'appréciation des différentes recettes sont différentes selon la méthode employée. Les résultats du test en salle montrent qu'il serait risqué de tenter de remplacer la recette originale par une des deux recettes diminuées en sucre. Cependant, les résultats du test à domicile ne confirment pas cette conclusion. Bien que le classement des recettes soit le même, la différence d'appréciation entre les recettes A et B n'est pas significative. La note de la recette C reste quant à elle bien inférieure aux deux autres comme dans le cas du test en salle. L'interaction significative *produit* \times *méthode* [$F_{(2,996)} = 25,92$; $p = 0,007$] donnée par l'ANOVA confirme le fait que la discrimination des produits obtenue par les deux tests n'est pas identique.

II.2.2 - Influence du contexte sur la notation de l'intensité sucrée par rapport à l'idéal

La figure 11 présente les distributions des réponses des sujets selon la méthode ainsi que les résultats des tests du χ^2 pour chaque recette.

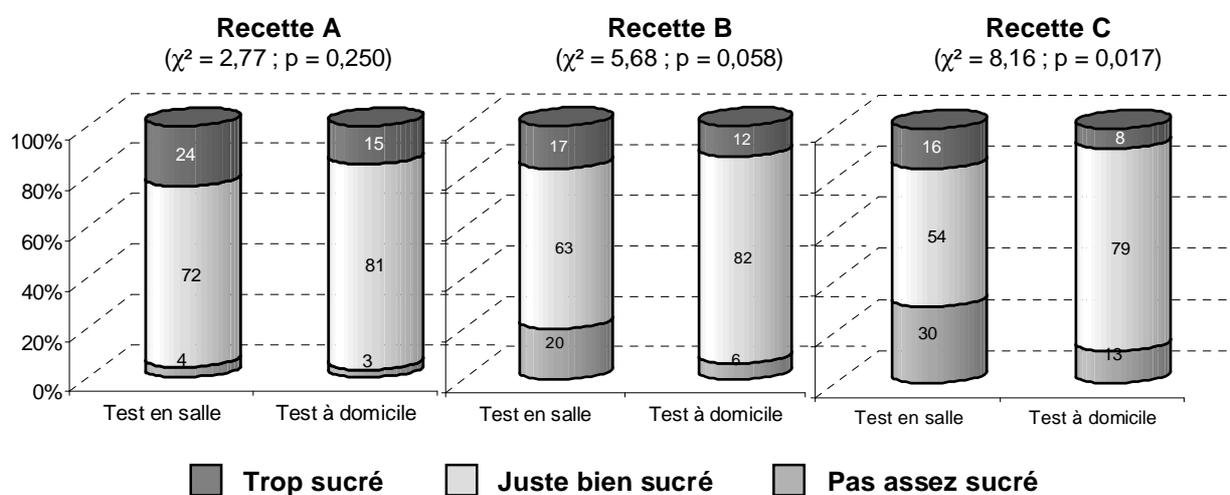


Figure 11 : Répartition des réponses des sujets concernant l'appréciation de l'intensité sucrée des trois recettes par rapport à l'idéal.

Nous remarquons que lors du test en salle, les deux recettes B et C ont été perçues comme pas assez sucrées par un pourcentage non négligeable de participants. Ce pourcentage de réponse diminue nettement lors du test à domicile pour les deux recettes et nous observons alors une amélioration de la proportion de sujets déclarants que les recettes sont « juste bien sucrées ». En revanche, la proportion de réponses « juste bien sucré » évolue peu pour la recette standard A entre les deux méthodes. Ces évolutions des réponses entre le test en salle

et le test à domicile sont en accord avec l'augmentation à domicile des notes hédoniques données aux recettes B et C et avec la relative stabilité de la note hédonique donnée à la recette A.

II.3 - DISCUSSION

Le premier résultat de cette étude concerne encore une fois la sous-notation des produits lors du test en salle par rapport au test à domicile. Les interprétations concernant ce phénomène ayant été détaillées dans l'article précédent, nous ne les re-citerons pas.

En ce qui concerne l'impact des conditions de test sur le niveau de discrimination des produits, l'interprétation précédente basée sur le mode de consommation des produits ne semble pas adaptée. En effet, les données de consommations (carnet de bord) obtenues grâce au test à domicile ont montré un profil de consommation assez proche de celui des laits fermentés à boire. Ainsi, les occasions de consommation se répartissent entre le matin au petit déjeuner (32%) et le reste de la journée. Les sujets ont, pour 75% des occasions de consommation, dégusté le produit en solitaire et ne l'ont accompagné d'aliments que dans 45% des cas. Nous avons pu voir dans l'étude sur les laits fermentés que ces situations de consommations sont finalement assez proches de ce que peut expérimenter un sujet lors d'un test en salle. Nous pouvons donc supposer que le mode de consommation du produit n'est pas en cause dans la modification des résultats entre le test en salle et le test à domicile. La littérature regorgeant de travaux ayant observé un impact de la quantité de produit dégusté sur l'appréciation de produits plus ou moins sucrés, nous pensons donc que la différence de quantité consommée entre le test en salle (quelques gorgées) et le test à domicile (1L minimum) pourrait expliquer cette différence de résultats entre les deux tests. En effet, d'après Lucas et Bellisle (1987) il semblerait que, spontanément, les consommateurs dirigent leur préférence vers les produits les plus sucrés lors de la dégustation d'une quantité faible de produit. Une consommation répétée d'une plus grande quantité de produit entraînerait un déplacement de cette préférence pour les produits les plus sucrés vers des produits moins sucrés (Chung et Vickers, 2007 ; Drewnowski *et al.*, 1987 ; Vickers *et al.*, 2001). De la même façon, Vickers et Holton (1998) montrent que lors d'une exposition ponctuelle à un produit, les consommateurs ont tendance à se diriger vers les produits ayant les plus fortes caractéristiques sensorielles. Lors de consommations répétées, cet attrait peut finalement devenir un défaut suite à un écoëurement ou à des effets post-ingestifs négatifs. La teneur en sucre des trois recettes étant la principale caractéristique sensorielle les différenciant, nous

pouvons supposer que ce phénomène de sous-notation des produits peu sucrés par rapport aux produits plus sucrés lors d'un test bref est celui que nous observons sur notre jeu de données et plus particulièrement pour la recette B. Alors que la recette B est automatiquement refoulée lors du test en salle classique par rapport à la recette standard après une faible dégustation, sa consommation sur du plus long terme à domicile en fait finalement un produit autant apprécié que la recette standard.

L'analyse de la relation entre la quantité globale consommée lors du test à domicile et l'appréciation des recettes (figure 12) montre que plus les sujets ont consommé une quantité importante du produit, plus ils lui ont donné une note d'appréciation élevée. Ceci peut rendre compte de l'utilisation de la variable « quantité consommée » comme indicateur du niveau d'appréciation des produits. Cependant, cette relation positive entre la note d'appréciation globale et la quantité consommée est particulièrement forte pour la recette B qui est peu appréciée par les sujets ayant consommé une seule bouteille alors qu'elle est autant appréciée que la recette A par les sujets ayant consommé deux bouteilles. Ainsi, cette interaction *produit* \times *quantité* pourrait illustrer le fait que la quantité consommée de la recette B influence positivement son appréciation.

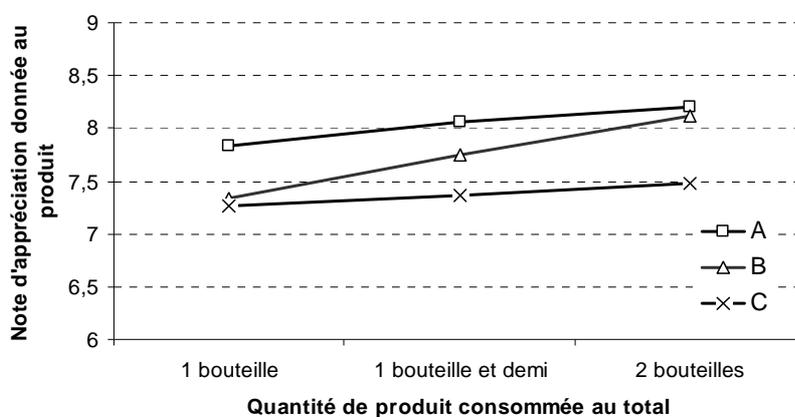


Figure 12 : Notes moyennes d'appréciation du produit selon la quantité globale consommée lors du test à domicile pour chacune des recettes

Cette observation met donc en évidence qu'outre le mode de consommation des produits, l'expérimentateur doit également être vigilant au fait que, lorsque des produits diffèrent fortement sur une seule caractéristique sensorielle comme le sucré, un contact prolongé entre les participants et les produits est nécessaire avant de récolter leur opinion.

III – TEST EN SALLE vs. TEST A DOMICILE : DES RESULTATS DIFFERENTS EN FONCTION DU SEXE DES SUJETS

Les travaux de King *et al.* (2007) ont mis en évidence que le sexe et l'âge des interviewés peuvent avoir une influence sur l'effet des conditions d'évaluation. Dans cette optique, il nous a paru intéressant de tester sur l'ensemble de nos études le rôle des caractéristiques socio-démographiques de la cible interrogée sur l'influence du contexte. Les études présentées précédemment n'ont pas mis en évidence d'influence des caractéristiques socio-démographiques des sujets sur les conclusions concernant la comparaison des tests en salle et à domicile. Une autre étude réalisée avec deux recettes de thé glacé a mis en évidence un impact du sexe des interviewés sur l'influence du contexte. Nous allons présenter les résultats de cette étude en abordant l'analyse des données par groupe de participants.

III.1 - MATERIELS ET METHODES

III.1.1 - Les produits

L'objectif de ces tests consommateurs est de positionner deux recettes concurrentes A et B d'une boisson au thé aromatisée à la pêche. Les profils sensoriels réalisés sur les deux marques montrent que les deux recettes sont significativement différentes en goût et en arôme mais également en aspect. Alors que la recette A est principalement une boisson au thé avec une légère aromatisation à la pêche, la recette B est plutôt une boisson à la pêche avec des notes d'aromatisation au thé.

III.1.2 - Les sujets

Au total, 491 personnes ont dégusté un des deux produits selon les conditions d'un test en salle ou d'un test à domicile. Les sujets devaient avoir consommé du thé glacé au moins une fois dans les six derniers mois. Chaque sujet n'a dégusté qu'une seule recette. Les effectifs pour chaque recette sont les suivants :

- test en salle : recette A = 122 sujets, recette B = 123 sujets,
- test à domicile : recette A = 125 sujets, recette B = 121 sujets.

La répartition des sujets selon leur âge et leur sexe est la suivante dans chaque cellule :

- 50% d'hommes et 50% de femmes,
- 50% de personnes entre 25 et 39 ans et 50% de personnes entre 40 et 65 ans.

III.1.3 - Les protocoles

Le test en salle

Les sujets sont recrutés dans les rues de trois villes différentes et invités à participer à une séance de dégustation. L'enquêteur présente au sujet un verre de boisson à la température de sortie du réfrigérateur. Le sujet boit la quantité de produit qu'il souhaite avant de lui donner une note d'appréciation globale de 1 à 10. L'enquêteur questionne ensuite le sujet sur quelques caractéristiques spécifiques du produit que nous ne développerons pas ici.

Le test à domicile

Le test à domicile consiste à déposer deux bouteilles de 2 litres de l'une ou l'autre des boissons accompagnées du questionnaire d'appréciation globale ainsi que d'un carnet de bord de consommation. Le consommateur doit alors consommer personnellement au moins 5 verres de la boisson et consigner les conditions dans lesquelles il les consomme avant de donner son opinion concernant le produit. Au bout d'une semaine, l'enquêteur revient au domicile du consommateur et récupère le questionnaire d'appréciation globale ainsi que les carnets de consommation puis questionne le sujet sur les caractéristiques spécifiques du produit.

III.1.4 - Analyse des données

Pour chacune des méthodes, les notes moyennes d'appréciation globale ainsi que les écart-types pour les deux recettes sont calculés. La significativité de la différence d'appréciation entre les deux produits est ensuite mesurée par un test de Student bilatéral non apparié.

Une ANOVA ($note = produit + méthode + âge + sexe + produit \times méthode + produit \times âge + produit \times sexe + méthode \times âge + méthode \times sexe + âge \times sexe + produit \times méthode \times âge + produit \times méthode \times sexe + produit \times sexe \times âge + méthode \times âge \times sexe$) est réalisée de manière à déterminer l'influence du protocole de test sur l'appréciation globale et l'impact des caractéristiques individuelles des sujets interrogés. Pour l'ensemble des analyses, le seuil de significativité a été choisi à 5%.

III.2 - RESULTATS**III.2.1 - Analyse des données de l'ensemble du panel**

Le tableau 3 présente les notes moyennes d'appréciation globale obtenues pour chacune des recettes en salle et à domicile.

Tableau 3 : Notes moyennes d'appréciation globale pour chaque recette selon la méthode et résultats des tests de Student respectifs.

	Test en salle		Test à domicile	
	Recette A	Recette B	Recette A	Recette B
Effectif	122	123	125	121
Note moyenne	6,93 ^b	7,50 ^a	6,98 ^b	7,64 ^a
Ecart-type	1,75	1,50	1,99	1,64
Test T (p-value)	t = 2,71 (p=0,008)		t = 2,80 (p=0,005)	

Quelle que soit la méthode, la boisson B est significativement plus appréciée que la boisson A. Les résultats de l'ANOVA (tableau 4) confirment la stabilité de cette conclusion étant donné que l'interaction *produit* × *méthode* n'est pas significative. De manière plus inattendue, nous observons pour la première fois que le niveau de notation entre les deux méthodes reste le même. En effet, nous n'assistons pas cette fois-ci à une sous-notation des produits lors du test en salle par rapport au test à domicile et l'effet *méthode* de l'ANOVA n'est pas significatif.

Tableau 4 : Résultats de l'ANOVA globale pour tester l'influence de la méthode

Effets	DDL	F ratio	p-value
Produit	1	15,57	<0,001
<i>Méthode</i>	1	0,38	0,536
<i>Age</i>	1	0,36	0,551
<i>Sexe</i>	1	0,00	0,978
<i>Produit</i> × <i>méthode</i>	1	0,13	0,718
<i>Produit</i> × <i>âge</i>	1	0,12	0,732
<i>Produit</i> × <i>sexe</i>	1	1,44	0,230
<i>Méthode</i> × <i>âge</i>	1	0,99	0,320
<i>Méthode</i> × <i>sexe</i>	1	1,04	0,309
<i>Age</i> × <i>sexe</i>	1	0,01	0,938
<i>Produit</i> × <i>méthode</i> × <i>âge</i>	1	0,83	0,363
Produit × <i>méthode</i> × <i>sexe</i>	1	9,89	0,002
Produit × <i>sexe</i> × <i>âge</i>	1	4,62	0,032
<i>Méthode</i> × <i>âge</i> × <i>sexe</i>	1	3,61	0,058

Au vu de ces résultats, nous serions tentés de conclure que le risque d'obtenir une information hédonique erronée en testant ce type de produits par l'intermédiaire d'un test en salle à la place d'un test à domicile est faible. Cependant, les résultats de l'ANOVA (tableau 4) mettent en évidence un effet significatif du sexe des interviewés sur l'interaction *produit* × *méthode*. Nous avons donc exploré ce jeu de données sur la base des évaluations données séparément par les hommes et par les femmes. Cependant, compte tenu du faible effectif de ces deux groupes de sujets, les conclusions obtenues sur ces deux jeux de données seront à prendre avec précaution⁵. Par ailleurs, même si nous nous intéressons dans ce chapitre uniquement à l'influence des caractéristiques individuelles des sujets sur les effets de contexte, nous pouvons remarquer que, quelle que soit la méthode, le sexe combiné à l'âge des interviewés influence les différences d'appréciation entre les deux recettes (interaction *produit* × *sexe* × *âge* significative).

III.2.2 - Analyse des données générées par les hommes

Le tableau 5 présente les notes moyennes d'appréciation globale données par les hommes aux deux produits selon les conditions d'évaluation. Contrairement aux résultats obtenus sur l'ensemble du panel, les conclusions des tests sont influencées par la méthode employée. Lors des conditions de dégustation du test en salle, les hommes apprécient plus fortement la boisson B alors que cette différence d'appréciation n'est plus significative lors du test à domicile. L'ANOVA (*note* = *produit* + *méthode* + *produit* × *méthode*) réalisée sur ce jeu de données met en évidence un effet *méthode* non significatif [$F_{(1,241)} = 1,30$; $p = 0,255$] ce qui confirme que les niveaux de notations entre le test en salle et le test à domicile ne sont pas différents. Par ailleurs, l'interaction *produit* × *méthode* significative [$F_{(1,241)} = 4,47$; $p = 0,035$] illustre que la différence d'appréciation des deux produits est bien influencée par la méthode.

Tableau 5 : Notes moyennes d'appréciation des deux recettes pour chaque méthode et résultats des tests de Student associés (hommes)

	Test en salle		Test à domicile	
	Recette A	Recette B	Recette A	Recette B
Effectifs	61	61	63	60
Note moyenne	6,52	7,77	7,21	7,57
Ecart type	1,79	1,21	1,87	1,59
Test T	t = 4,48		t = 1,15	
(p-value)	(p < 0,001)		(p = 0,252)	

⁵ Cette différence d'effet de contexte selon le sexe des participants a cependant été confirmée lors de la répétition de cette étude avec la présentation des produits non masqués (résultats non présentés).

III.2.3 - Analyse des données générées par les femmes

Les moyennes des notes d'appréciation globale données aux produits selon la méthode sont répertoriées dans le tableau 6.

Tableau 6 : Notes moyennes d'appréciation des deux recettes pour chaque méthode et résultats des tests de Student associés (femmes)

	Test en salle		Test à domicile	
	Recette A	Recette B	Recette A	Recette B
Effectif	61	62	62	61
Note moyenne	7,34	7,23	6,76	7,70
Ecart-type	1,61	1,70	2,09	1,70
Test T (p-value)	t = 0,40 (p=0,690)		t = 2,75 (p=0,007)	

Nous obtenons un profil de résultats très différent de celui généré par les hommes. En effet, même si les conclusions statistiques des deux méthodes sont différentes comme dans le cas du jeu de données des hommes, cette fois-ci le niveau de discrimination des produits est plus important lors du test à domicile que lors du test en salle. Alors que le test en salle ne montre pas de différence significative entre l'appréciation des deux recettes, la recette B est plus appréciée que la recette A lors des dégustations à domicile. Les résultats de l'ANOVA ($note = produit + méthode + produit \times méthode$) confirment la significativité de l'influence de la méthode sur le niveau de discrimination des produits [interaction $produit \times méthode$: $F_{(1,242)} = 5,48$; $p = 0,020$]. L'influence de la méthode sur le niveau de notation reste, quant à elle, non significative [effet $méthode$: $F_{(1,242)} = 0,05$; $p = 0,814$].

III.3 - DISCUSSION

Cette dernière comparaison de protocoles révèle de nouvelles informations quant à l'influence du contexte de dégustation sur l'appréciation des produits.

Pour la première fois, nous n'avons pas observé de différence de notation entre les deux protocoles de dégustation. Ceci pourrait provenir du fait que les conditions du test en salle étaient particulièrement favorables. En effet, ce test a eu lieu au moment d'un épisode de canicule pendant lequel les températures extérieures avoisinaient les 35°C. Ainsi, le recrutement des sujets pour déguster une boisson fraîche a été particulièrement facilité. La dégustation des boissons fraîches en salle était donc particulièrement appropriée à l'état de soif probablement élevé des participants et l'état d'humeur global des interviewés en a peut-être été affecté conduisant à des notes d'appréciation relativement similaires au domicile.

La deuxième information apportée par cette étude concerne un aspect des tests consommateurs que nous avons peu abordé jusque-là : l'importance des sous-groupes de sujets lors de la mise en place des tests consommateurs. En effet, nous remarquons ici que l'analyse sur l'ensemble des données n'apporte pas les mêmes informations que l'analyse par groupe de sujets en ce qui concerne :

- les conclusions statistiques par rapport à l'appréciation relative des deux recettes,
- l'influence du contexte de dégustation sur les résultats.

Si nous nous étions contentés d'analyser les données d'un point de vue global, nous aurions conclu que les conditions de test n'influencent pas la différence d'appréciation des produits. Cette conclusion aurait d'ailleurs été supportée par l'analyse des données de consommation du test à domicile qui mettent en évidence un profil de consommation plutôt de type nomade. En effet, les occasions de consommation sont réparties tout au long de la journée avec un maximum de 25% dans l'après-midi en dehors du goûter. Les consommateurs étaient seuls dans 60% des cas et n'ont accompagné leur boisson d'aliment que dans 20% des occasions de consommation. Une fois encore ces conditions de consommation se rapprochent donc fortement de celles du test en salle. Cependant, l'analyse des données par sous-groupe de sujets a finalement mis en évidence que, bien au contraire, les conditions d'évaluation (salle ou domicile) influencent significativement les conclusions des tests. Etant donné que les deux sous-groupes aboutissent à des résultats inverses, l'observation des données globales ne permet pas de mettre en évidence cette influence du contexte sur les écarts d'appréciation des produits.

Il n'est pas si étonnant que les goûts des hommes et des femmes pour cette catégorie de produits ne soient pas les mêmes. En effet de nombreuses études consommateurs réalisées en interne chez Danone ainsi que dans la littérature (Monneuse *et al.*, 1991b) ont déjà montré que les hommes et les femmes n'apprécient pas forcément les mêmes produits. Il est en revanche plus surprenant que les conditions d'évaluation n'influencent pas les résultats de la même façon pour les hommes et les femmes. La figure 13 illustre l'influence du sexe des interviewés sur l'impact du contexte (interaction *produit* \times *méthode* \times *sexe*). Ainsi, lors du test à domicile, les femmes présentent finalement une appréciation plus élevée pour le produit B que pour le produit A. Le comportement des hommes est, quant à lui, inverse et finalement, un contact prolongé avec le produit à domicile diminue la différence d'appréciation entre les deux recettes jusqu'à la rendre non significative. Les femmes discriminent donc mieux les produits à domicile alors que les hommes les discriminent mieux en salle.

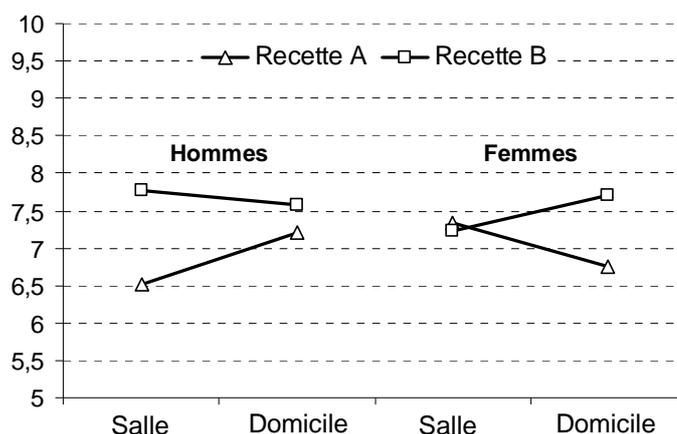


Figure 13 : Illustration de l'interaction produit \times méthode \times sexe

Face à ces résultats, la seule explication plausible que nous avons envisagée concerne le fait que les hommes et les femmes n'aient pas consommé les produits dans les mêmes conditions lors du test à domicile. Cependant, l'exploration des données des carnets de consommation à domicile n'a pas démontré de différence très nette entre les conditions de consommation des hommes et des femmes. Les hommes et les femmes ont consommé les produits globalement aux mêmes moments, à la même température et dans un environnement alimentaire proche. En revanche, nous avons observé de légères différences en ce qui concerne l'environnement social de la dégustation (les hommes étaient plus souvent seuls que les femmes lorsqu'ils ont dégusté les produits). La quantité de boisson consommée est également différente : les hommes ont bu une quantité supérieure de boisson que les femmes pendant la durée du test à domicile. Cette différence de quantité consommée à domicile pourrait éventuellement être à la base de ces résultats contradictoires entre les hommes et les femmes concernant l'influence du contexte.

Par ailleurs, le piètre masquage des bouteilles à domicile a également pu intervenir. Alors que le test en salle est réalisé avec les produits dégustés en aveugle, le contrôle de l'anonymat des produits distribués à domicile dans leur emballage initial masqué est quant à lui moins certain. Ainsi, cette comparaison du test en salle et du test à domicile pourrait aussi bien refléter l'influence de la marque ou de l'emballage du produit sur son appréciation que l'influence du contexte globale de dégustation. Nous pourrions ainsi supposer que les images véhiculées par les marques reconnues des produits testés peuvent ne pas avoir eu le même impact sur l'appréciation des recettes pour les hommes et pour les femmes.

Même si nous avons des difficultés à formuler une interprétation solide de cette influence du sexe des interviewés sur l'impact du contexte, cette étude pointe du doigt, une fois de plus, la complexité de l'anticipation de l'influence du contexte sur les résultats des tests consommateurs.

Tableau 7: Résultats des cinq comparaisons de tests en salle et à domicile

	Lait fermenté à boire		Biscuit feuilleté au fromage		Eau gazeuse		Boisson à base de jus de fruit et de lait		Thé glacé			
	SALLE	DOM	SALLE	DOM	SALLE	DOM	SALLE	DOM	Hommes		Femmes	
Produits	2 recettes Danone: - standard (A) - allégé en sucre et en matière grasse (B)		2 marques concurrentes: - Danone (A) - Marque de distributeur (B)		2 marques concurrentes: - Danone (A) - Principal concurrent (B)		3 recettes Danone: - standard (A) - -25% de sucre ajouté (B) - sans sucre ajouté (C)		2 marques concurrentes: - Danone (A) - Principal concurrent (B)			
Méthode	SALLE	DOM	SALLE	DOM	SALLE	DOM	SALLE	DOM	SALLE	DOM	SALLE	DOM
Monadique séquentiel												
Effectifs	239 sujets	241 sujets	240 sujets	240 sujets	161 sujets	160 sujets	<i>Test non réalisé</i>		<i>Test non réalisé</i>			
Note A	6,96	7,24	7,14	7,91	6,74	7,54						
Note B	6,55	6,95	7,02	7,09	6,20	7,14						
Résultat statistique	A > B (p=0,004)	A > B (p=0,029)	A = B (p=0,448)	A > B (p<0,001)	A > B (p=0,007)	A > B (p=0,033)						
Effet méthode	** (p=0,006)		** (p<0,001)		** (p<0,001)							
Effet produit × méthode	ns (p=0,627)		** (p=0,003)		ns (p=0,639)							
Monadique pur												
Effectifs	environ 120 sujets par produit		environ 120 sujets par produit		environ 80 sujets par produit		environ 160 sujets par produit		environ 60 sujets par produit		environ 60 sujets par produit	
Note A	7,22	7,17	7,51	8,07	6,80	7,63	7,70	8,07	6,52	7,21	7,34	6,76
Note B	7,15	7,10	7,42	7,47	6,33	7,33	6,81	7,80	7,77	7,57	7,23	7,70
Note C							6,28	7,37				
Résultat statistique	A = B (p=0,782)	A = B (p=0,743)	A = B (p=0,703)	A > B (p=0,001)	A = B (p=0,123)	A = B (p=0,265)	A > B > C (p<0,001)	A = B > C (p<0,001)	A < B (p<0,001)	A = B (p=0,252)	A = B (p=0,690)	A < B (p=0,007)
Effet méthode	ns (p=0,777)		* (p=0,031)		** (p<0,001)		** (p<0,001)		ns (p=0,255)		ns (p=0,814)	
Effet produit × méthode	ns (p=0,975)		* (p=0,070)		ns (p=0,680)		** (p=0,007)		* (p=0,035)		* (p=0,020)	

IV - TEST EN SALLE vs. TEST A DOMICILE : DISCUSSION GENERALE

Le tableau 7 présente les résultats obtenus pour les cinq comparaisons de protocoles réalisées. Les trois premières expérimentations présentent les résultats pour les données obtenues à la fois en monadique séquentiel et en monadique pur. Les deux dernières expérimentations ont été menées uniquement en monadique pur. Les résultats de la dernière étude sont présentés par sous-groupe de sujets pour lesquels l'influence des conditions de test sur les résultats est différente. Pour les tests effectués en monadique séquentiel, les moyennes d'appréciation des produits sont comparées par une ANOVA ($note = sujet + produit + ordre$) et la significativité de la différence est donnée par l'effet *produit*. Pour les tests effectués en monadique pur sur deux produits, un test de Student non apparié permet de tester la différence d'appréciation. Les tests effectués en monadique pur sur trois produits sont analysés par une ANOVA ($note = produit$). Afin de faciliter la lecture des conclusions de chaque test nous avons symbolisé la non-significativité d'une différence d'appréciation entre deux recettes par un symbole d'égalité entre les deux recettes même si les écarts d'appréciations n'ont pas été statistiquement testés en ce qui concerne leur égalité. Enfin, pour chaque comparaison de méthodes, les résultats de l'ANOVA ($note = produit + méthode + produit \times méthode$) permettent de quantifier l'influence des conditions de test sur l'appréciation des produits testés.

Nous allons à présent synthétiser les informations essentielles que l'ensemble de ces comparaisons méthodologiques nous a permis d'obtenir en ce qui concerne l'influence du contexte d'évaluation :

- sur le niveau de notation des produits,
- sur les écarts d'appréciation des produits (les conclusions statistiques).

Nous essaierons de dégager à partir de ces résultats des critères nous permettant d'anticiper les conséquences de l'utilisation d'un test en salle au lieu d'un test à domicile.

IV.1 - IMPACT DES CONDITIONS D'EVALUATION SUR LE NIVEAU DE NOTATION DES PRODUITS

IV.1.1 - Hypothèses interprétatives

Une première constatation concerne l'existence d'un effet *méthode* significatif dans la majorité des comparaisons avec un niveau de notation de l'appréciation des produits plus élevé à domicile qu'en salle. Nous observons cependant quelques exceptions, notamment pour

l'étude sur les thés glacés pour laquelle le niveau de notation des deux recettes est identique en salle et à domicile. L'existence de cet effet *méthode* est en accord avec ceux d'autres études de comparaison des tests en situation naturelle et en situation artificielle. Nous avons élaboré plusieurs hypothèses pour expliquer ce phénomène de sous-notation des produits en salle par rapport au test à domicile.

Parmi ces hypothèses, l'une concerne les conditions d'évaluation à domicile a priori plus confortables que les conditions d'un test en salle. En effet, les salles de test sont plutôt peu décorées et ne sont pas toujours très confortables ce qui peut donc jouer sur le bien-être global des sujets au moment de l'évaluation du produit. Au contraire, l'évaluation à domicile est supposée être effectuée dans un environnement plus confortable, plus familier ce qui peut donc favoriser les notations quel que soit le produit évalué par rapport à une situation moins confortable. L'étude d'Edwards *et al.* (2003) a effectivement montré une influence du prestige du lieu de test sur les notes hédoniques données à un même plat. Cependant, peu d'auteurs ont étudié spécifiquement l'impact du confort du lieu de test sur la mesure hédonique et ceux qui l'ont fait (Bonin *et al.*, 2001 ; King *et al.*, 2004) n'ont alors pas observé réellement de modification des notes hédoniques suite à une amélioration du confort de la situation de test (un accueil agréable, du matériel haut de gamme, des tables décorées). Cette hypothèse de l'impact du confort du lieu sur le niveau de notation reste donc à explorer.

Une hypothèse peut-être plus probable concerne le fait que la participation du sujet à une séance de dégustation en salle est généralement réalisée au détriment d'une autre activité prévue initialement. Cette contrainte peut rendre les interviewés plus stressés que dans le cas d'une situation de consommation à domicile vécue sans stress lié au temps de l'évaluation. Ainsi, l'état d'humeur et le bien-être global du participant en salle seraient moins favorables ce qui l'entraînerait à être plus sévère envers les produits. Les conditions d'évaluation très formelles en salle peuvent également entraîner le sujet à percevoir la séance de dégustation comme un examen et le pousser à être beaucoup plus exigeant envers les produits qu'il ne l'aurait été dans un contexte plus naturel de consommation.

Ces hypothèses liées au manque de confort des salles de test et à l'aspect formel des séances de dégustation sous-entendent donc que, quel que soit le produit testé, les sujets ont tendance à avoir des références de notation plus basses en salle qu'à domicile même si un produit est globalement apprécié de la même façon. Ce phénomène de sous-notation ne devrait donc pas être dépendant des produits testés. Cependant, certains jeux de données que nous avons récoltés n'ont pas mis en évidence cette différence de notation. Par ailleurs, si

nous observons les écarts entre les moyennes des notes (tous produits confondus) obtenues en salle et à domicile pour les études ayant mis en évidence un effet *méthode* significatif (tableau 8), nous remarquons que cet écart de note n'est pas identique pour toutes les études que nous avons menées. En effet, les écarts de notation pour les laits fermentés et les biscuits sont beaucoup moins importants que ceux observés dans le cas des eaux gazeuses et des boissons à base de jus de fruit et de lait. Ainsi, même si nous pouvons affirmer que la majorité des produits seront sous-évalués lors d'un test en salle par rapport à un test à domicile, il nous est en revanche impossible de quantifier cet écart de notes.

Tableau 8 : Ecart entre les moyennes des notes obtenues lors du test en salle et du test à domicile pour les jeux de données présentant un effet méthode significatif.

Etude	Données	Ecart de notes (domicile - salle)
Laits fermentés à boire	Monadique séquentiel	0,34
Biscuits feuilletés au fromage	Monadique pur Monadique séquentiel	0,42 0,31
Eaux gazeuses	Monadique pur Monadique séquentiel	0,91 0,87
Boisson à base de jus de fruit et lait	Monadique pur	0,82

Cette instabilité de l'écart de notes entre les deux protocoles selon les études remet donc en cause l'hypothèse d'un simple décalage de notation lié à l'état interne du sujet et indépendant des produits. Ceci suggère donc l'intervention de caractéristiques liées aux produits testés dans l'élaboration des hypothèses pour expliquer ces différences de notation entre le test en salle et le test à domicile. Les conditions de test plus ou moins favorables à la consommation du produit testé pourraient être en cause dans cette différence du niveau de notation. En effet, le test à domicile permet au sujet de consommer le produit au moment où il en a envie et dans des conditions appropriées et optimales. Ainsi, le seul fait de permettre les conditions naturelles de consommation du produit favoriserait le plaisir de consommation du participant et donc son opinion sur le produit testé. L'augmentation des notes hédoniques à domicile suite à un contact prolongé avec les produits pourrait également être le reflet d'un effet dit de simple exposition (Zajonc, 1968) dû à la familiarisation des sujets avec les produits testés. Cet effet est cependant peu probable car les produits testés dans ces expérimentations sont des catégories de produit relativement banalisées et les sujets recrutés étaient consommateurs de la catégorie de produits testée.

IV.1.2 – Implications concernant l'interprétation de la modification de l'appréciation d'un produit

Cette différence du niveau de notation des produits en salle et à domicile soulève une difficulté concernant l'interprétation de la modification de la note hédonique obtenue par un seul produit d'une méthode à l'autre en termes de modification de son appréciation. Prenons comme exemple les données obtenues en monadique séquentiel pour les biscuits feuilletés au fromage lors du test en salle et lors du test à domicile. La figure 14 représente les moyennes des notes données aux deux produits pour chacune des méthodes.

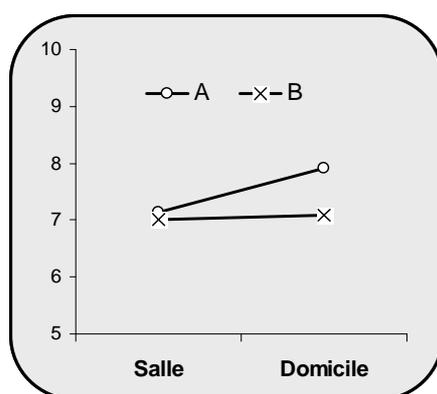


Figure 14 : Scores moyens d'appréciation globale observés pour l'étude sur les biscuits feuilletés au fromage

Au vu de ces résultats, nous pouvons interpréter l'influence des conditions de dégustation sur l'appréciation de chaque recette de différentes façons. Si nous nous basons sur les données observées sans prendre en compte la différence du niveau de notation, nous pouvons conclure que le produit B est apprécié de la même façon en salle et à domicile alors que la performance du produit A semble s'améliorer lors du test à domicile. Cependant, dans l'hypothèse que la méthodologie du test à domicile entraîne intrinsèquement des notes plus élevées qu'un test en salle pour un même niveau réel d'appréciation, le fait que la note hédonique de la recette B soit identique entre les deux méthodologies peut être interprété comme une diminution de son appréciation à domicile. L'augmentation de la note hédonique du produit A illustrerait alors simplement le décalage de notation entre le test en salle et le test à domicile pour une appréciation équivalente.

Ces données peuvent donc aboutir à deux interprétations distinctes en termes d'évolution de l'appréciation des deux recettes selon que l'on tienne compte ou non du décalage de notes :

- Interprétation sans le décalage de notation : l'appréciation de la recette A augmente à domicile alors que l'appréciation de la recette B est identique. Une caractéristique

positive de la recette A aurait donc été particulièrement appréciée lors du test à domicile mais pas forcément lors du test en salle.

- Interprétation avec le décalage de notation : l'appréciation de la recette A reste stable alors que l'appréciation de la recette B diminue à domicile. La recette B présenterait donc un défaut qui aurait influencé négativement son appréciation lors de la consommation à domicile mais pas lors de l'évaluation en salle.

Les données des questions diagnostiques obtenues selon les deux protocoles de test ont été analysées afin d'essayer d'interpréter l'appréciation plus élevée pour la recette A que pour la recette B observée lors du test à domicile mais pas lors du test en salle. Les réponses à la question diagnostique sur l'intensité salée des produits mettent en évidence une faiblesse de la recette B à la fois en salle et à domicile. En effet, celle-ci est perçue comme « trop salée » par plus de 30% des sujets alors que la recette A n'est perçue comme « trop salée » que par environ 10% des participants. Les corrélations entre les réponses données à chaque question diagnostique et la note d'appréciation globale ont été ensuite analysées séparément pour les données du test en salle et du test à domicile. Nous avons alors remarqué que la contribution des réponses à la question de l'intensité salée des biscuits sur l'appréciation globale de la recette B est beaucoup plus importante dans le cas du test à domicile que dans le cas du test en salle (figure 15). Il semblerait donc que cette caractéristique « trop salée » de la recette B ait bien été perçue dans les deux conditions de test mais qu'elle ait influencé négativement l'appréciation globale de la recette B uniquement lors du test à domicile.

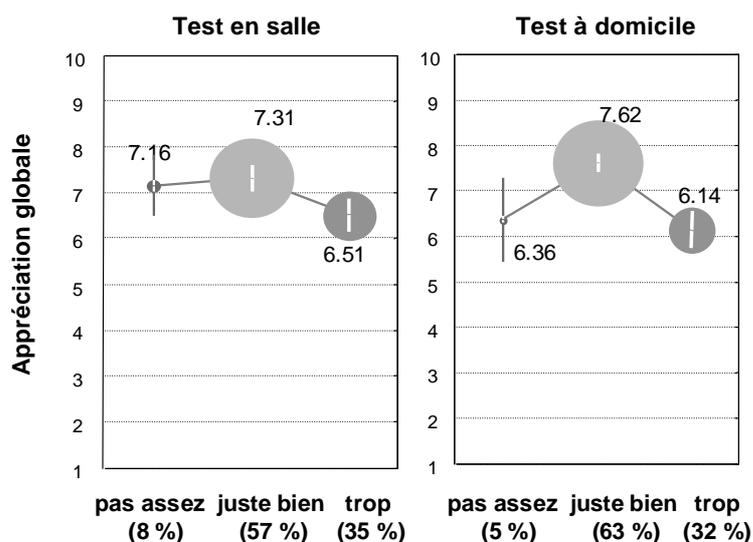


Figure 15 : Score moyen d'appréciation de la recette B en fonction des réponses données à la question JAR sur l'intensité salée pour le test en salle et le test à domicile.

Cette ambiguïté dans l'interprétation de la modification du jugement hédonique d'une méthode d'évaluation à l'autre est particulièrement importante dans notre travail de comparaison de méthodes puisqu'elle complique l'interprétation de l'évolution des notes hédoniques de chaque produit testé séparément. C'est pourquoi, afin de ne pas aboutir à des conclusions et des interprétations hasardeuses, nous avons choisi de discuter uniquement de l'évolution de la différence de note entre les produits comparés et non de l'évolution de chacune des notes des produits pris séparément les uns des autres.

IV.2 - IMPACT DES CONDITIONS D'ÉVALUATION SUR LES CONCLUSIONS STATISTIQUES DES TESTS

IV.2.1 - Instabilité de l'influence du contexte selon les études : les conditions d'évaluation n'influencent pas systématiquement les conclusions des tests

Les comparaisons méthodologiques réalisées n'ont pas toutes mis en évidence une influence du contexte sur les écarts de notation entre les produits comparés et in fine sur les conclusions des tests. Nous pouvons formuler plusieurs hypothèses liées aux produits testés concernant l'existence ou l'absence d'influence des conditions de dégustation sur les conclusions hédoniques obtenues.

IV.2.1.1 - Influence du contexte selon le mode de consommation de la catégorie de produits

L'analyse des données d'usage obtenues sur les trois catégories de produits étudiées dans le premier article (boissons au lait fermenté, biscuits feuilletés au fromage et eaux gazeuses) nous a permis d'intégrer le mode de consommation des différentes catégories de produit dans nos hypothèses explicatives concernant la variabilité de l'influence des conditions d'évaluation sur les conclusions des tests. En effet, le mode de consommation des laits fermentés à boire, proche des conditions de consommations imposées lors d'un test en salle, pourrait expliquer que les conditions d'évaluation n'ont pas influencé l'écart d'appréciation entre les deux échantillons testés. La différence d'appréciation des deux biscuits feuilletés au fromage est en revanche sensible aux conditions de tests et s'expliquerait par le fait que la consommation de ce produit est très contextualisée et s'organise le plus souvent autour de l'apéritif pendant le test à domicile. La catégorie des eaux gazeuses est plus problématique. En effet, 50% des occasions de consommation de ces eaux gazeuses sont réalisées au cours d'un repas, c'est-à-dire à un moment bien spécifique. Les autres occasions

de consommations sont plutôt nomades et étalées tout au long de la journée. Ainsi, ces occasions de consommation nomades des produits pendant le test à domicile peuvent expliquer les résultats identiques à ceux du test en salle malgré la présence d'occasions de consommation pendant un repas. Le tableau 9 synthétise les caractéristiques du mode de consommation des produits qui seraient susceptibles d'entraîner une différence de résultats entre un test en salle et un test à domicile.

Tableau 9 : Caractéristiques du mode de consommation des produits qui pourraient entraîner des résultats identiques ou différents entre un test en salle et un test à domicile

Modalités des variables		
Quantité consommée par consommation	Faible : conditionnement individuel	Importante
Durée du moment de consommation	Rapide	Long
Moment de consommation	Non spécifique : tout au long de la journée	Spécifique : matin, midi, soirée..
Occasion de consommation	Non spécifique	Spécifique : fête, pique nique, apéritif...
Environnement social	Consommation en solitaire	Consommation avec d'autres personnes
Environnement alimentaire	Aucun autre aliment, ni boisson consommés	Combinaison d'aliments et/ou de boissons
Préparation du produit	Non personnalisée	Personnalisée : température de l'aliment, assaisonnement...
Manipulation du produit	Standardisée : toujours dans l'emballage ou toujours hors de l'emballage	Personnalisée : consommation au verre ou à la bouteille par exemple
Test en salle vs. Test à domicile	=	≠

Cette grille a pour objectif d'aider les expérimentateurs à évaluer les risques qu'ils prennent lorsqu'ils mettent en place un test en salle au lieu d'un test à domicile sur une catégorie de produits bien spécifique. Elle sert donc de mise en garde pour les expérimentateurs mais permet également de les aider à décider de la possibilité de mettre en place un test en salle au lieu d'un test à domicile. Ainsi, si l'usage du produit testé répond à plusieurs critères présents dans la colonne de gauche (ce qui est le cas des laits fermentés à boire), l'expérimentateur a la possibilité de mettre en place un test en salle en prenant moins de risque d'aboutir à des résultats différents d'un test à domicile que si l'usage du produit testé s'inscrit plutôt dans la colonne de droite. Il est évidemment important de préciser que cette grille n'est pas exclusive et que ce n'est pas parce qu'une seule des caractéristiques

présente dans la colonne de droite s'applique au mode de consommation d'un produit que l'utilisation d'un test en salle est à bannir définitivement. De la même manière, nous avons pu voir que pour les deux dernières expérimentations, les profils d'usage des deux catégories de produits étaient plutôt favorables à l'utilisation d'un test en salle ce qui ne nous a pas empêché d'observer des différences de résultat entre le test en salle et le test à domicile. L'utilisation d'une telle grille de recommandations est donc à manier avec précautions.

IV.2.1.2 - Influence du contexte selon d'autres caractéristiques liées aux produits testés

Outre le mode de consommation des produits, d'autres caractéristiques intrinsèques aux produits comparés peuvent être mises en cause dans l'obtention d'un effet de contexte. En effet, la nature de la différence sensorielle entre les produits peut être source de différents résultats entre un test en salle et un test à domicile. Il semble que si cette différence est basée sur un seul critère comme le sucre, le gras ou le sel, les tests en salle ne permettent pas de révéler correctement les performances des produits. En effet, les faibles quantités de produit consommées lors d'un test en salle entraînent probablement moins d'effets de saturation que les quantités plus élevées consommées lors d'un test à domicile. Par ailleurs, lors des tests à domicile, la consommation d'une grande quantité de produits présentant des teneurs élevées en gras, sucre ou sel entraînent probablement la mise en place de croyances négatives associées à ces caractéristiques nutritionnelles des produits. Enfin, même si nous n'avons pas pu illustrer l'importance des effets post-ingestifs dans la formation de l'appréciation d'un produit, il convient également de faire l'hypothèse que des produits possédant des bénéfices ou inconvénients liés à leur consommation seront également plus susceptibles d'aboutir à des résultats différents entre un test en salle ponctuel et un test à domicile durant lequel les individus ont la possibilité de ressentir ces effets.

Nous pouvons donc compléter la grille précédente (tableau 10) en ajoutant les deux variables dont nous venons de discuter dans les caractéristiques à prendre en compte lors du choix d'une méthode.

Tableau 10 : Caractéristiques concernant les produits à prendre également en compte lors du choix d'une méthode

Modalités des variables		
<i>Nature de la différence entre les produits comparés</i>	Multi-attributs	Uni-attribut (sucré, salé, gras...)
<i>Effets post-ingestifs</i>	Faibles	Importants
Test en salle vs. Test à domicile	=	≠

IV.2.2 - Stabilité de l'influence du contexte selon la procédure de présentation des produits

Grâce aux trois premières expérimentations mises en place sur un grand nombre de sujets, nous avons pu analyser l'influence de la méthode employée à la fois en monadique pur et en monadique séquentiel.

Naturellement, nous pensions que les comparaisons de méthodes en monadique séquentiel auraient plus de chance de mettre en évidence une différence de résultats que les comparaisons de méthodes réalisées en monadique pur. En effet, outre les conditions de dégustation qui sont modifiées de la même façon en monadique pur et en monadique séquentiel, la procédure en monadique séquentiel fait intervenir une modification supplémentaire entre le test en salle et le test à domicile. Les données obtenues en monadique séquentiel en salle et à domicile sont influencées par l'environnement de dégustation mais également par le délai entre les évaluations des deux produits. Alors que les évaluations en salle sont espacées de quelques minutes, les évaluations à domicile sont en général espacées de plusieurs jours. Ainsi, l'impact de l'évaluation du premier produit sur le deuxième produit en salle est sûrement différent lors du test en salle et lors du test à domicile. Dans une toute autre considération, le nombre plus élevé de données en monadique séquentiel aurait également pu jouer en faveur de la démonstration d'un effet significatif de la méthode sur les résultats comparé aux données obtenues en monadique pur. Malgré ces hypothèses, nous n'avons pas observé d'impact plus prononcé du contexte en monadique séquentiel qu'en monadique pur. En effet, même si les résultats en monadique séquentiel et en monadique pur ne sont pas toujours équivalents lors de la même expérimentation, l'impact de la méthode sur l'appréciation globale, lui, reste identique.

IV.2.3 - Description de l'influence du contexte sur les conclusions statistiques des tests

Nous venons de voir que les conclusions statistiques des tests ne sont pas systématiquement influencées par les conditions d'évaluation. Il convient cependant de s'interroger sur la possibilité d'anticiper la nature de la manifestation de ces effets de contexte.

L'existence de l'influence du contexte de dégustation sur les résultats est problématique principalement quant elle s'applique aux écarts d'appréciation entre les produits dégustés. L'effet le plus embarrassant consisterait en une modification du classement de l'appréciation des produits selon les conditions de test mais nos expérimentations n'ont jamais mis en

évidence une influence aussi forte des conditions de test ce qui est donc plutôt rassurant. En revanche, nous avons observé des modifications concernant le niveau de discrimination des différents produits ce qui peut donc aboutir à des prises de décisions différentes. Reprenons par exemple les résultats de l'étude sur la possibilité de diminuer la teneur en sucre de la boisson à base de jus de fruit et de lait. Alors que les résultats du test en salle condamnent définitivement la recette avec 25% de sucre ajouté en moins, les données du test à domicile en font un bon candidat pour remplacer la recette standard. L'autre résultat particulièrement intéressant qui bouscule légèrement les a priori concerne le fait qu'une méthode n'est pas plus discriminante qu'une autre. En effet, le test en salle aboutit à des résultats plus discriminants que le test à domicile pour les boissons à base de jus de fruit et de lait. L'étude sur les biscuits salés aboutit quant à elle au résultat inverse avec le test à domicile plus discriminant. Plus surprenant encore, les niveaux de discrimination du test en salle et du test à domicile varient pour l'étude sur les thés glacés en fonction de la cible interrogée : les hommes sont plus discriminants en salle et les femmes sont plus discriminantes à domicile. Ainsi, le critère de « meilleure discrimination » du test en salle qui est souvent annoncé ne se retrouve pas dans nos expérimentations.

Le bilan concernant l'influence de la méthode sur les conclusions des tests est proche de ce que nous avons dégagé de la bibliographie sur les comparaisons des tests en situation artificielle de consommation et des tests en situation naturelle. Il est déjà compliqué de pouvoir anticiper si des écarts de notes seront sensibles ou non aux conditions de dégustation. Il est encore plus difficile de dégager un consensus de nos résultats qui nous permettrait de pouvoir anticiper l'évolution du pouvoir discriminant des tests suite à l'influence du contexte de dégustation.

Ce premier travail de comparaison du test en salle et du test à domicile nous a surtout permis de bien prendre conscience de l'influence des conditions de test sur les notes hédoniques collectées au moyen de nos protocoles traditionnels. Nous avons également pu soulever la complexité de la nature des effets de contexte et notre incapacité à les identifier clairement. L'interprétation des données de ces comparaisons de protocoles est donc essentiellement basée sur des hypothèses qui ont le mérite d'introduire une discussion sur la pertinence des protocoles des tests en salle.

Chapitre II

PROPOSITIONS METHODOLOGIQUES

A – PROBLEMATIQUE INDUSTRIELLE**I – TEST EN SALLE vs. TEST A DOMICILE : « CONFLICTING DESIDERATA »****I.1 – TEST EN SALLE**

Dans le premier chapitre de ce document, nous avons présenté une analyse du problème de l'utilisation des tests en salle en conditions artificielles de dégustation pour accéder à l'appréciation des aliments. Nous avons ainsi pu observer, à la fois dans la littérature et à travers nos expérimentations, que les conditions de consommation artificielles des tests en salle peuvent entraîner des résultats différents du test à domicile concernant la comparaison de l'appréciation de différentes recettes. Les décisions stratégiques basées sur des résultats de tests en salle peuvent donc ne pas être optimales et entraîner, dans le cadre de problématiques de développement de formules, des conséquences économiques importantes pour l'entreprise.

En dépit de ce problème de validité externe, le test en salle présente de nombreux avantages qui expliquent son succès auprès des industriels. En effet, aujourd'hui, en France, 70% des tests consommateurs sur des produits alimentaires sont réalisés en salle (source Repères). Le test en salle est effectivement facile à mettre en place dans un délai relativement court et avec un budget raisonnable. Par ailleurs, le contrôle des conditions d'évaluation permet de récolter des données plus précises et plus faciles à interpréter que celles du test à domicile. Tuorila et Lähteenmäki (1992) considèrent que « les situations en laboratoire sont peut-être artificielles mais que c'est un contexte pour étudier le comportement alimentaire qui nous permet de contrôler des variables qui sont souvent confondues entre elles lors des études en situation naturelle ». L'article 5 présenté en annexe compare la robustesse de la différence d'appréciation entre les deux recettes de lait fermenté obtenue avec le test en salle et avec le test à domicile face à une diminution de l'effectif du panel interrogé. Les résultats de cette comparaison ont mis en évidence une légère supériorité de la robustesse de la conclusion du test en salle par rapport à celle du test à domicile qui pourrait être attribuée au meilleur contrôle des conditions de dégustation en salle qu'à domicile.

I.2 TEST A DOMICILE

Bien que le caractère standardisé et artificiel des conditions dans lesquelles les évaluations hédoniques sont recueillies lors des tests en salle tend à être remis en cause, les inconvénients des tests en situations naturelles sont également reconnus (Meiselman, 1992 ;

Mela *et al.*, 1992 ; Pliner, 1992 ; Rolls et Shide, 1992). Comme le précisent Rolls et Shide (1992), « le problème avec les tests en situation naturelle est que la plupart des méthodes ne sont pas précises, ne permettent pas de manipulations expérimentales et sont très coûteuses ». Les limites budgétaires et logistiques du test à domicile freinent donc aujourd'hui son utilisation en routine par les industriels.

Si nous nous basons sur des devis de test réalisés pour les produits Danone, les coûts proposés pour un test à domicile en France sont généralement 50% plus élevés que les coûts proposés pour un test en salle. Cette différence de budget provient principalement du pré-recrutement nécessaire des sujets, de leur rémunération et du déplacement des enquêteurs à domicile. Par ailleurs, la quantité de produits à fournir pour le test est beaucoup plus importante compte tenu du fait que chaque consommateur doit disposer d'une quantité de produits équivalente à une semaine de consommation ce qui participe donc à augmenter le coût d'un test à domicile par rapport à un test en salle.

En ce qui concerne la logistique, de nombreux aspects du test à domicile rendent parfois sa mise en œuvre très difficile voire impossible. La principale difficulté du test à domicile réside dans la réalisation de l'anonymat des produits à déguster. Lors des tests en salle, l'anonymat des produits est généralement réalisé par un simple masquage de l'emballage initial (pour des doses individuelles) ou par la présentation d'un prélèvement du produit. Cependant, dans le cas du test à domicile, la manipulation des produits par les sujets empêche un simple masquage par des boîtes ou des autocollants faciles à enlever. Ainsi, les tests à domicile nécessitent souvent un re-conditionnement total des produits dans un emballage neutre. Malheureusement, de nombreux produits ne peuvent se soumettre à une telle manipulation sans que leurs propriétés organoleptiques n'en soient modifiées. C'est notamment le cas des boissons gazeuses ou encore des produits laitiers ayant une texture ferme. Ces difficultés expliquent donc que le test à domicile est peu utilisé en aveugle lors de tests sur des produits déjà commercialisés. L'évaluation de produits prototypes pose, quant à elle, moins de problème puisque les développeurs peuvent choisir l'emballage des produits. L'autre problème soulevé par le test à domicile concerne les produits à faible durée de conservation. En effet, la durée d'exposition à chaque produit testé d'une semaine ainsi que la rotation de l'ordre de présentation oblige fréquemment les expérimentateurs à faire tester des échantillons de lots de production différents ce qui peut être problématique lorsque la variabilité de production est importante. Le test à domicile présente également un problème d'efficacité dû au fait qu'il ne peut pas être mis en place sur un ensemble de plus de deux ou trois produits successifs compte tenu du délai entre chaque produit évalué. Le dernier

inconvéniént majeur du test à domicile concerne la faible réactivité de cette méthode. En effet, sa mise en place et sa réalisation nécessitent en général un délai d'au moins un mois (pour deux produits) avant d'obtenir les premiers résultats alors qu'un test en salle peut être organisé de manière à ce que les premiers résultats soient disponibles en une semaine. Après avoir énuméré l'ensemble des problèmes liés à la mise en place d'un test à domicile, il nous est plus facile de comprendre pourquoi les professionnels des tests consommateurs utilisent plus souvent les tests en salle. Même si les expérimentateurs étaient prêts à payer plus cher pour garantir une meilleure validité prédictive de leurs résultats, de lourdes contraintes logistiques les en empêchent.

I.3 - CONFLICTING DESIDERATA

Les tableaux 11 et 12 présentent les raisons de préférer le test en salle et celles de préférer le test à domicile relevées dans les ouvrages consacrés aux tests consommateurs (Lawless et Heymann, 1999 ; Meilgaard *et al.*, 1991 ; Stone et Sidel, 1993). Nous observons dans ces tableaux que les avantages d'une méthode font généralement échos aux limites de l'autre méthode. Ainsi, les conditions standardisées du test en salle en font une méthode pratique, peu coûteuse et permettant d'obtenir une mesure hédonique précise et facile à interpréter bien que potentiellement peu prédictive de la réalité. Bien que plus coûteuse, la méthode du test à domicile avec ses conditions plus naturelles de dégustation permet, quant à elle, d'obtenir une mesure hédonique potentiellement plus prédictive de la réalité mais peu contrôlée et donc plus difficile à interpréter en termes actionnables pour le développement. Cette opposition entre la précision du test en salle et la validité supposée du test à domicile nommé *conflicting desiderata* par Brinberg et McGrath (1985) est donc au cœur du problème de ce travail de thèse.

Tableau 11 : Les raisons de préférer le test en salle

AVANTAGES DU TEST EN SALLE	LIMITES DU TEST A DOMICILE
Coût moins élevé	Coût plus élevé
Obtention des résultats rapidement	Difficile et long à mettre en œuvre
Possibilité de réaliser les tests dans une caravane mobile pour changer facilement de lieux et entrer en contact rapidement avec des consommateurs différents	Pré-recrutement nécessaire de consommateurs sur des bases de données qui sont donc moins "naïfs" que ceux recrutés ponctuellement dans la rue
Conditions de contrôle relativement bonnes (environnement, préparation et présentation des produits, compréhension des instructions...): données précises et facile à interpréter	Pas de contrôle de l'environnement : grande variabilité de réponses dûe à la variabilité des conditions de consommation
Respect des instructions Peu de données manquantes	Nécessite un contrôle qualité important pour vérifier le respect des instructions (est ce que le produit a bien été consommé par les bonnes personnes ?) Possibilité élevée de données manquantes
Souvent plus aisé pour les approches comparatives : possibilité de tester plusieurs produits durant une séance	Evaluation de beaucoup de produits impossible

Tableau 12 : Les raisons de préférer le test à domicile

AVANTAGES DU TEST A DOMICILE	LIMITES DU TEST EN SALLE
Le produit est préparé et consommé dans des conditions normales d'utilisation	Le sujet est en dehors de son contexte habituel de consommation : préparation, heure de dégustation, environnement social et alimentaire, etc
L'appréciation obtenue est basée sur une consommation répétée d'une quantité importante du produit	Les opinions sont recueillies après un contact bref avec le produit
De nombreuses informations sur l'utilisation du produit peuvent être obtenues et être utilisées dans l'interprétation de l'information hédonique	Les informations sur les usages et attitudes recueillies sont uniquement déclaratives et peuvent donc ne pas correspondre à ce que le consommateur expérimente vraiment chez lui
Pas de biais dus à la dégustation de plusieurs produits dans un laps de temps limité: comparaison implicite fiable avec le produit habituel du sujet	Perturbation des références du sujet qui peut conduire à des jugements non pertinents

II - OBJECTIF

Etant donné l'importance des budgets dépensés par les industriels pour accéder à l'opinion des consommateurs par l'intermédiaire de tests en salle, nous pouvons imaginer que la perspective d'obtenir des résultats erronés les contrarient. Plus préoccupant encore, les conséquences économiques engendrées par les décisions stratégiques (lancement, modification ou retrait d'une formule) prises sur la base de ces mauvais indicateurs d'appréciation justifient indiscutablement ce travail de thèse.

Nous avons précédemment remis en question à la fois la validité externe des données obtenues dans les conditions artificielles de dégustation des tests en salle ainsi que la précision des données du test à domicile due au manque de contrôle des conditions d'évaluation. Aujourd'hui, les commanditaires de tests consommateurs sont donc en attente de solutions méthodologiques faciles à mettre en place et leur permettant d'améliorer le pouvoir prédictif de leurs données hédoniques tout en conservant un degré de contrôle suffisant sur les conditions d'évaluation.

La deuxième partie de ce travail de thèse a donc pour objectif de proposer des solutions méthodologiques pour améliorer à la fois la pertinence et la précision des données hédoniques collectées lors d'un test organoleptique tout en considérant les contraintes budgétaires et logistiques incompressibles du monde industriel. En effet, tout projet d'amélioration d'une méthode dans un souci d'optimisation de la performance des résultats obtenus prend très généralement la direction opposée de la praticité de la méthode. Même si les expérimentateurs désirent trouver un moyen de minimiser les risques d'obtenir une information hédonique déconnectée de la réalité, les contraintes du monde industriel sont telles que toutes les solutions ne sont pas envisageables. Ainsi, les propositions méthodologiques explorées doivent pouvoir être mises en place en routine dans un laps de temps et avec un budget raisonnables tout en permettant un contrôle des conditions de test. Compte tenu de ces contraintes, nous avons fait le choix de travailler sur l'amélioration de la validité prédictive du test en salle. La deuxième partie expérimentale de ce travail de thèse concerne donc une réflexion méthodologique basée sur des propositions d'améliorations de différents aspects des protocoles de test consommateurs en salle.

III - CHOIX DES AXES METHODOLOGIQUES EXPLORES

Nous avons sondé la littérature au sujet de protocoles de test en salle qui diffèrent de la simple question d'appréciation ou de préférence posée après la consommation des produits en conditions artificielles. Ce panorama nous a permis de mettre en évidence deux grands axes de travail envisageables pour améliorer la validité externe de l'information hédonique récoltée tout en permettant un contrôle des conditions de test :

- les protocoles de test en salle qui intègrent les conditions habituelles de consommation des produits (les approches orientées « situation »),
- les protocoles de test en salle qui mettent en œuvre des mesures d'appréciation indirectes en vue de favoriser l'implication du participant dans ses réponses (les approches orientées « implication »).

La figure 16 présente donc les deux types d'approches que nous avons explorées dans l'objectif d'améliorer le pouvoir prédictif des données hédoniques collectées lors d'un test en salle. Nous observons sur cette figure que chacune des voies d'amélioration a fait l'objet de deux protocoles basés sur des méthodes relevées dans la bibliographie. Nous allons brièvement introduire le potentiel de chacune de ces méthodes.

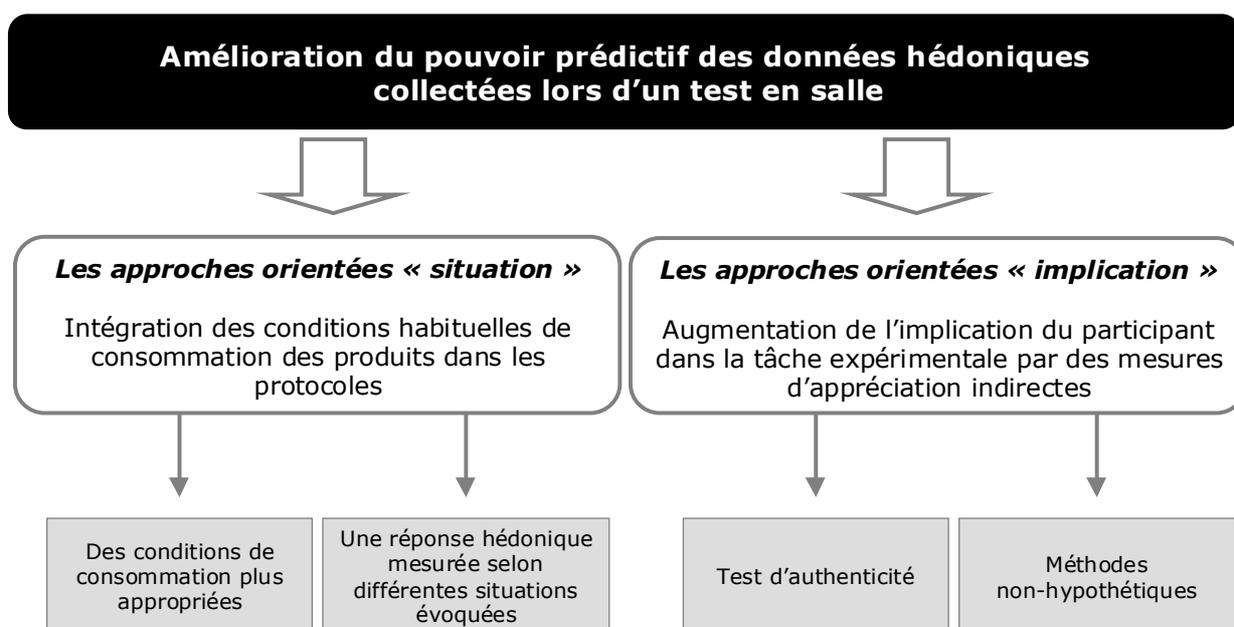


Figure 16 : Récapitulatif des différentes approches explorées pour améliorer la prédiction des données hédoniques obtenues lors d'un test en salle

III.1 – LES APPROCHES ORIENTEES « SITUATION »

La revue de la littérature de ce domaine nous a permis de sélectionner deux solutions pour intégrer les conditions habituelles de dégustation dans la mesure hédonique :

- proposer un environnement physique de consommation plus approprié à l'évaluation du type de produit considéré,
- récolter une information de nature hédonique qui soit fonction des situations de consommation du produit.

III.1.1 - Des conditions physiques de consommation plus appropriées

Certains auteurs ont essayé d'améliorer les conditions physiques de dégustation des tests en salle de façon à les rendre plus appropriées à la consommation habituelle du produit. Il s'agit par exemple du protocole de Posri *et al.* (2001) qui permet un ajout de lait et de sucre dans des échantillons de thés ou encore de celui de De Graaf *et al.* (2005) qui permet au sujet de choisir l'aliment testé. King *et al.* (2004) ont également tenté de modifier de manière séquentielle le statut de quelques variables de contexte de façon à rendre les conditions de dégustation de plus en plus naturelles (repas, choix des produits, environnement social...). Mais finalement, peu d'études proposent réellement des protocoles de test qui rendraient globalement la situation de consommation plus naturelle, plus appropriée. L'analyse en temps et en fréquence développée par Köster (1981) abonde dans ce sens et propose de travailler sur des méthodes d'observation de la quantité de produit consommée lors d'évènements pendant lesquels les sujets consomment librement les produits testés.

Etudier les possibilités techniques de mise en place de ce type de tests proposant des environnements physiques globalement plus appropriés à la consommation des produits est une des solutions méthodologiques que nous avons souhaité explorer.

III.1.2 - Une réponse hédonique mesurée selon différentes situations de consommation évoquées

D'autres chercheurs se sont intéressés de manière plus interprétative à la variation de la mesure hédonique selon les conditions de consommation. En effet, utiliser l'appréciation globale des produits comme variable explicative du comportement de consommation serait insuffisant pour certains auteurs (Kamenetzky *et al.*, 1957 ; Pilgrim et Kamen, 1963 ; Sidel *et al.*, 1972). Une personne peut effectivement apprécier un aliment sans pour autant le trouver approprié pour un lieu ou une situation donnée. Ainsi, nous assistons au développement de protocoles de tests hédoniques intégrant le rôle de la situation d'utilisation (de consommation)

sur l'appréciation des produits. Ces méthodes consistent à récolter des notes d'à-propos des produits pour différentes situations d'usage. La mesure d'« *appropriateness* » combinée à une note hédonique d'appréciation globale est maintenant fréquemment utilisée pour étudier l'appréciation de produits multi-usages (Cardello et Schutz, 1996 ; Lähteenmäki et Tuorila, 1997, 1998)

Cette mesure de l'à-propos de produits de catégories très différentes pour plusieurs situations de consommation semble bénéfique puisqu'elle permet de prendre en compte que les attentes des consommateurs varient selon les situations de consommation. Cependant, l'information apportée par cette mesure dans le cas d'échantillons très proches appartenant à la même catégorie de produits semble plus limitée et son intérêt reste à démontrer. Il nous a donc semblé intéressant de tester ce type de protocole dans le cadre de la comparaison de produits proches sensoriellement. Köster (1981b) souligne également que la limite de cette méthode réside dans les moyens mis en œuvre pour évoquer les différentes situations. Nous devons donc réfléchir à des moyens permettant d'aider le sujet à se plonger réellement dans ses propres situations quotidiennes.

III.2 – LES APPROCHES ORIENTEES « IMPLICATION »

Plutôt que de travailler sur les conditions de consommation des produits, d'autres auteurs se sont penchés sur la validité prédictive des mesures hédoniques traditionnelles comme les notes ou les rangs d'appréciation. En effet, ce questionnaire direct d'appréciation ou de préférence aboutit vraisemblablement à des réponses peu impliquées de la part des sujets et donc particulièrement sensibles à l'influence du contexte artificiel dans lequel le jugement est émis.

Aujourd'hui, la pratique des tests consommateurs en salle a mis en évidence une difficulté concernant le manque d'attention et d'implication des participants. En effet, les sujets sont généralement recrutés ponctuellement et ont donc très peu de temps à accorder à l'enquêteur. Cet état d'esprit général a alors tendance à entraîner les sujets à donner des scores hédoniques plutôt proches du hasard et donc peu discriminants. Dans le domaine des enquêtes, ce biais est appelé « *satisficing bias* » (Simon, 1956). Il correspond à la tendance des participants à mettre en œuvre un processus de traitement des informations moins rigoureux et efficace qu'il ne l'aurait été dans des conditions réalistes. Cette stratégie permet en effet à l'individu d'aboutir rapidement à un résultat qui est assez bon mais qui n'est pas

forcément optimal. En ce qui concerne les réponses données dans des enquêtes, le « *satisficing* » se manifesterait par :

- le choix de l’option explicitement offerte de réponse « sans opinion »,
- le choix des réponses socialement souhaitables,
- le manque de discrimination entre de multiples objets d’intérêts,
- le biais d’acquiescement qui est la tendance à être d’accord avec n’importe quelle affirmation indépendamment de son contenu.

Esposito et Jobe (1991) et Krosnick (1991) précisent que cette tendance au « *satisficing* » est particulièrement influencée par les conditions de test lors d’enquêtes. D’après Stolte (1994), les conditions favorables au « *satisficing* » sont globalement des conditions favorables à la distraction du sujet (un lieu agité, pas de rémunération des sujets et très peu de temps pour répondre au questionnaire) alors que les conditions défavorables impliquent des distractions minimales pour les sujets (un lieu calme, rémunération des participants, autant de temps qu’il souhaite pour répondre au questionnaire). Malheureusement, ces conditions favorables au « *satisficing* » ressemblent fortement aux conditions qu’expérimentent les sujets qui acceptent de participer aux tests hédoniques en salle tels qu’ils sont pratiqués actuellement. Outre des conditions de test plus calmes, la motivation élevée du participant est également citée comme permettant de diminuer la manifestation du « *satisficing* ». Dans le cadre des tests hédoniques en salle, il semble difficile de diminuer drastiquement les conditions de distraction des sujets. En revanche, nous pouvons essayer d’améliorer la motivation du participant par le biais de la tâche expérimentale qui lui est demandée. L’acte d’achat du produit est un acte qui véhicule des notions de satisfaction et de regret pour le consommateur c’est-à-dire des conséquences cognitives liées à l’achat. Or, cette notion de risque, composante importante de l’implication, n’intervient pas au niveau de la déclaration d’une note d’appréciation ou d’une préférence. Le manque d’implication des participants dans leurs réponses lors des tests consommateurs pourrait donc être une des causes logiques du manque de prédiction des données déclaratives. Par ailleurs, une implication plus importante de la part du participant dans une enquête permettrait éventuellement de s’affranchir en partie de l’influence du contexte à partir du moment où les réponses hédoniques ont des conséquences concrètes pour le sujet.

Nous avons donc exploré deux types de méthodes qui proposent des tâches dont les réponses véhiculent une forte implication du participant et qui sont donc supposées mieux refléter ses « vraies » préférences : le test d’authenticité et les méthodes non-hypothétiques.

III.2.1 - Le test d'authenticité

La méthode du « test d'authenticité » développée par Mojet et Köster (1986) consiste à demander au sujet le niveau d'authenticité d'un produit suite à un amorçage émotionnel négatif envers les produits. Le type de question indirecte ainsi que l'enjeu du test pousse les sujets à répondre de manière très affective, spontanée et sans contrôle rationnel comme il semble que ce soit le cas lors de l'émission d'un jugement en situation naturelle. Cette méthode s'est montrée particulièrement efficace pour révéler des différences fines d'appréciation de produits comme de la bière par des consommateurs réguliers de la marque testée. Cependant, l'apport de cette méthode mise en place avec des sujets moins familiers des produits pour lesquels, par ailleurs, la notion d'authenticité a moins de sens reste à étudier.

III.2.2 - Les méthodes non-hypothétiques

Le manque d'implication des participants lors de l'émission d'une attitude concerne le manque de conséquences concrètes pour le sujet, c'est-à-dire une prise de risque nulle. Ainsi, des méthodes de test s'attachent à impliquer le consommateur dans ses réponses par le biais des conséquences directes de ses réponses. Le choix d'un produit à emporter à la fin du test est parfois mis en place pour accéder aux préférences des consommateurs (Leon *et al.*, 1999 ; Levy et Köster, 1999). Cependant, cette mesure comportementale est critiquable compte tenu du fait que les choix ne sont pas toujours réalisés en accord avec les préférences du sujet. En effet, un produit peut être choisi par le consommateur pour un autre membre de sa famille. La curiosité d'emporter le produit le plus nouveau, le plus étrange peut également faire basculer le participant vers un choix différent de sa préférence.

Les méthodes non-hypothétiques dérivées de l'économie expérimentale comme les méthodes d'enchères permettent d'intégrer une notion de risque dans le choix d'un produit à emporter. Le bénéfice de ces méthodes est qu'elles introduisent tout simplement du réalisme dans les protocoles de test en faisant réellement acheter les produits par les sujets. Ces méthodes non-hypothétiques basées sur l'évaluation de l'intensité du désir de posséder réellement le produit testé commencent à voir le jour depuis une dizaine d'années dans les études consommateurs menées par les organismes de recherche (Lange *et al.*, 2002 ; Noussair *et al.*, 2004a ; Stefani *et al.*, 2006). Cependant, leur utilisation par les entreprises comme outil pour accéder aux préférences des consommateurs reste aujourd'hui nulle. Nous avons donc souhaité explorer la possibilité de mettre en place une telle méthode non-hypothétique dans le cadre d'un test en salle mené par un institut d'étude.

B – LES APPROCHES ORIENTEES « SITUATION »**I – DES CONDITIONS DE CONSOMMATION PLUS APPROPRIEES : LES TESTS SITUATIONNELS (ARTICLE 2)**

La première partie de ce travail de thèse sur la comparaison de données collectées lors d'un test en salle et lors d'un test à domicile nous a particulièrement sensibilisé à l'impact des conditions de dégustation artificielles sur l'appréciation des produits. Ainsi, il nous semblait tout à fait adéquat de débiter ce travail de propositions méthodologiques en essayant d'apporter un peu de « réalisme » dans les protocoles de test en salle. De cette façon nous avons tenté de répondre à l'appel de Meiselman (1992) qui milite pour recentrer l'étude du comportement alimentaire sur des « vraies personnes consommant de vrais aliments dans de vraies situations de consommation ». Cardello *et al.* (2000) ont déjà tenté l'expérience en proposant en laboratoire de tester des pizzas en situation de repas et des hors d'œuvres en situation sociale. Ils concluent leur expérience en soulevant la nécessité de développer des environnements de test « virtuels » dans lesquels les mesures attitudinales seraient obtenues dans des environnements qui simulent au mieux les situations réelles de consommation d'intérêt. D'autre part, Köster (2003) dénonce les conditions de dégustation dénuées de « significations » empêchant l'activation naturelle des événements passés concernant l'objet évalué et, dans ce sens, ajoute que les mesures en laboratoire sont complètement légitimes mais que de nouvelles méthodes doivent être développées pour mesurer les réponses hédoniques dans un contexte plus significatif, plus parlant pour le consommateur et donc plus naturel. Par ailleurs, il semblerait que le seul fait de manipuler quelques variables de contextes liées à l'expérimentation de manière à les rendre plus « naturelles » permettrait à d'autres variables de contexte internes liées au sujet de se manifester plus naturellement également.

Etant donné ces considérations, nous avons souhaité explorer la possibilité de proposer des environnements physiques de dégustation plus appropriés à la consommation des produits testés. Cette première proposition méthodologique a été testée dans le cadre de trois situations de consommation fréquemment associées aux produits commercialisés par Danone : un repas, un petit déjeuner et un événement social. Les résultats de ces tentatives de tests en salle dans des conditions de consommation plus réalistes (tests situationnels) ont été comparés aux résultats obtenus dans des tests en salle traditionnels et sont présentés dans l'article qui suit.

ARTICLE 2 : Boutrolle I., Köster E.P., Arranz D., Rogeaux M., Danzart M. Delarue J.

Some attempts to improve realism in Central Location Tests.

Food Quality and Preference. Article en préparation

RESUME

La mauvaise validité externe des résultats obtenus lors de tests consommateurs en salle est souvent attribuée aux conditions artificielles de dégustation des produits imposées par les expérimentateurs. Cette étude présente différentes tentatives de protocoles pour rendre les conditions de dégustation plus naturelles lors de la mise en place d'un test en salle. Trois situations de consommation spécifiques construites pour accroître l'à-propos de la situation de consommation selon le type de produit impliqué ont été mimées : une situation de repas pour l'évaluation d'eaux gazeuses, une situation de petit déjeuner pour des pains suédois et une situation sociale pour des crackers salés. Les résultats hédoniques obtenus pour chacun de ces tests « situationnels » ont été comparés à ceux obtenus dans des conditions artificielles de tests en salle. L'influence des conditions de dégustation des produits sur l'information hédonique récoltée a été observée pour chaque expérience. Le test en situation de repas ne montre pas de différence significative d'appréciation globale entre les deux marques d'eau gazeuse comparées alors que le test en salle traditionnel aboutit à une différence d'appréciation significative. En revanche, les réponses aux questions diagnostiques dans les deux conditions d'évaluation révèlent des différences entre les deux échantillons. L'étude sur les pains suédois illustre l'effet inverse. Les participants perçoivent les deux échantillons différemment selon les conditions d'évaluation. En revanche, ces différences de perception n'influencent pas les résultats de l'appréciation globale. Le test en situation sociale réalisé sur les deux marques de crackers montre que les données comportementales recueillies dans une situation naturelle de consommation aboutissent à des résultats différents des notes hédoniques recueillies dans les conditions traditionnelles d'un test en salle.

Mots-clés: *Situation de consommation, Test en salle, Appréciation, Questions diagnostiques, Effets de contexte.*

ABSTRACT

The artificial conditions of traditional Central Location Tests (CLT) are assumed to limit the external validity of hedonic results. The possibility to integrate more realism in the tasting conditions of a CLT under budget and logistical constraints was investigated. Three specific eating situations designed to enhance the appropriateness of the eating situation according to the product type involved were mimicked in a CLT setting: a meal situation for sparkling waters, a breakfast situation for crispy-breads and a social event situation for salted crackers. Hedonic output obtained in each of these situational CLTs was compared to that obtained in the traditional CLT conditions. Impact of the tasting conditions on hedonic output was observed in each experiment. The meal-CLT reduced the difference in overall liking between the two sparkling water brands that the traditional CLT revealed. Nevertheless, the diagnostic answers illustrated some differences between the two samples in both tasting conditions. The crispy breads study revealed the opposite effect. Here, diagnostic items revealed some very significant differences in the way the participant perceived the two crispy bread samples depending on the tasting conditions. However, these differences had no impact on the overall liking results. The social event experiment with the two crackers showed that behavioural data collected in a natural eating situation yielded different and more insightful results than hedonic scores collected under traditional CLT conditions.

Keywords: *Eating situation, Central location test, Liking, Diagnostic questions, Contextual effects.*

SOME ATTEMPTS TO IMPROVE REALISM IN CENTRAL LOCATION TESTS

Isabelle Boutrolle^{a,c}, Egon Peter Köster^b, Delphine Arranz^a, Marc Danzart^c, Julien Delarue^c

^a *Consumer science, Danone Research, RD 128, 91767 Palaiseau Cedex, France*

^b *Centre for Innovative Consumer Studies (CICS), Wageningen University Research Bornsesteeg 59, 6708PD
Wageningen, The Netherlands*

^c *Laboratoire de Perception Sensorielle et Sensométrie, AgroParisTech, 1 avenue des Olympiades, 91744 Massy
Cedex, France*

INTRODUCTION

In the early stages of the development of consumer test methodology, Schaefer (1979) promoted the use of central location (CLT) or laboratory tests under standardised tasting conditions in the assessment of food product acceptability. Compliance with instructions, manner of examining samples and ways of responding are easy to monitor and to control in this way. Such controlled consumer test methodologies were developed in order to minimise “noise” in the gathered data. At that time, “noise” referred to all the uncontrollable variables that were not related to the sensory food properties which could influence the food acceptability. Even if the precision of the data is a valuable reason for the choice of controlled protocols, we cannot overlook the fact that the artificial conditions reduce the predictive validity of the data. In fact, since the 1980’s, food acceptance and food choice have been modelled within a contextual framework. Thus, contextual factors appear in numerous schematic representations as determinants of food preferences (Khan, 1981; Randall & Sanjur, 1981) or food choice (Furst, Connors, Bisogni, Sobal, & Falk, 1996; Steenkamp, 1993; Williams, 1983). Meiselman (1996) proposed to define the context as the variables surrounding the food, the individual and the consumption situation. Rozin and Tuorila (1993) divided the context into factors affecting attitude and choice at the time of eating (simultaneous factors) and factors that exert their influence either prior or subsequent to the eating experience (temporaneous factors). Table 1 provides an overview of the contextual factors often cited in the literature for their influence on food acceptability measures (hedonic scores, preferences, choice or intake) and structured according to both the Meiselman (1996) and the Rozin and Tuorila (1993) classification.

Table 1. Examples of contextual factors cited in the literature for their influence on food acceptability.

	Food related factors	Individual related factors	Experimental related factors
Past factors	– Prior contacts with the tasted sample	– Food historic: episodic memory, habits	– Previous tasted sample
Simultaneous factors	– Amount of food consumed at each consumption occasion – Information concerning the product (price, origin, benefit information) – Preparation of the product (temperature, seasoning)	– Physiological state (hunger, thirst) – Psychological state (mood, motivation, reward)	Eating situation: – Time of the day – Social environment – Food environment – Comfort of the location – Choice of the tasted product – Physical effort to obtain the food
Future factors	– Anticipation of future physiological effects (satiety sensation, digestive discomfort)	– Anticipation of future events	– Next tasted sample

Many authors (Lawless & Heymann, 1999; Meiselman, 1992) point out the necessity to take real eating conditions into account in order to assess “real” food acceptability, i.e. to ensure good predictive validity of the gathered data. King, Weber, Meiselman, and Lv (2004) noticed that excluding these variables from research may oversimplify the participants’ eating experience, thus providing incomplete and, in some cases, misleading results. Accordingly, repeated in-home tasting under normal circumstances (HUT) is now largely advised. This allows the contextual variables to normally influence the eating experience and thus the hedonic judgement. However, this recommendation is still not applied often. HUT in fact presents numerous budget and logistical constraints such as time consumption or difficulty to provide blind samples. In addition to this, HUT methodology is often criticised because the tasting conditions are not well controlled. The variabilities in preparation method, in time of use and in combinations of the tested product with other foods lead to a large response variability that is difficult to interpret in terms suitable for product development.

Therefore, the call of Kanarek and Orthen-Gambill (1986) for “quasi-natural studies that capture the advantages of both the laboratory and the real world” is still up-to-date. However, with the available hedonic test protocols, it seems difficult to simultaneously optimise both precision (advantage of laboratory studies) and external validity (advantage of real world studies) of hedonic data. In fact, different ways to address this issue could be investigated. One could try to improve the precision of HUT data by imposing specific and homogeneous

tasting conditions by the way of precise instructions. However, compelling people to eat the tested food under certain circumstances would produce a considerable loss of the “naturalistic” feature brought by the HUT. Another trade-off between precision and validity would be the improvement of the appropriateness of CLT tasting conditions. To do this, it would be interesting to make the various controlled contextual factors brought by the CLT protocol more adapted to the consumption of the products. Improvement of appropriateness of the tasting conditions in CLT or lab test has been already attempted in previous studies. However, most of the studies investigated the manipulation of food or experimental related factors separately:

- sample serving size: small vs. large sample (Lähteenmäki & Tuorila, 1994, 1995; Popper, Maller, & Cardello, 1989),
- temperature of the food: appropriate vs. inappropriate (Cardello & Maller, 1982),
- preparation of the food: imposed vs. free (Matuszewska, Barylko-Pikielna, Szczecinska, & Radzanovska, 1997; Posri, MacFie, & Henson, 2001),
- time of the day: appropriate vs. inappropriate (Birch, Billman, & Richards, 1984; Kramer, Rock, & Engell, 1992),
- food environment: item tasted alone vs. food combination (King et al., 2004),
- social environment: self vs. social (Cardello, Schutz, Snow, & Leshner, 2000),
- possibility to choose the tested product: no choice vs. choice (De Graaf et al., 2005; King et al., 2004),
- comfort of the setting: non-cosy vs. cosy (Bonin, Chambres, & Bernard, 2001; King et al., 2004).

Few other studies investigated the appropriateness of the CLT tasting conditions in a global manner (Hersleth, Mevik, Naes, & Guinard, 2003; King, Meiselman, Hottenstein, Work, & Cronk, 2007). The topic of this paper is to explore this methodological approach and to show the impact of the combination of contextual factors when switching from artificial tasting conditions to more appropriate tasting conditions in CLT. According to De Graaf et al. (2005) and to Boutrolle, Delarue, Arranz, Rogeaux, and Köster (2007), CLT have an especially bad predictive performance when the products to be tested are strongly related to a specific consumption situation. Along this line of thinking, we chose to mimic three specific eating situations in a CLT setting: a breakfast, a meal and a social event. Checking the feasibility of implementing these kinds of situations in a CLT context under budget and logistical industrial constraints was a related objective. The three different studies were achieved in separate ways. The first experiment compared hedonic ratings of two sparkling waters collected either

in a classical CLT or in a CLT as part of a meal context. The second experiment compared hedonic ratings of two crispy breads collected either during a classical CLT or in a CLT in the form of a breakfast. In the last experiment, a social CLT situation with repeated consumption possibilities was achieved to assess the acceptability of two salted crackers. We explored whether behavioural data (amount consumed) gathered in the social context led to different hedonic results than data obtained with traditional CLT hedonic ratings.

1. EXPERIMENT 1: MEAL CONTEXT

1.1. MATERIALS AND METHODS

1.1.1. Subjects

Two different groups of women were recruited. One group (N = 161; 45.7 ± 11.3 years) participated in the traditional CLT design and the other group (N = 160; 47.4 ± 9.8 years) took part in the experimental meal CLT design. Only women who, during meals, usually drank sparkling water at room temperature were recruited.

1.1.2. Samples

Liking for two national brands of sparkling water (brand A and brand B) was assessed. According to the sensory profiles, brand B was more sparkling than brand A and otherwise the two brands were very similar. No information was given to the subjects regarding the products (brand, price...). The samples evaluated in both the traditional and the meal CLT came from the same product batches. The sparkling water samples were presented at ambient temperature (18-20°C).

1.1.3. Traditional CLT procedure

Participants were recruited in the street throughout the day (between 10:00 a.m. and 6:00 p.m.) and were invited to attend a 20 minutes tasting session. As it is usual in CLT, the rooms were set up for individual evaluations, avoiding social interaction between participants. Each subject was presented sequentially with the two samples in a balanced order. Subjects were instructed to drink as much as they wished, but at least half of the glass. Then, using self-administered questionnaires, they were asked to rate their overall liking for the sparkling water on a 10 point scale (1 = dislike very much; 10 = like very much). After having

evaluated the two samples for overall liking, respondents received the two products a second time, and were asked to rate them with regard to five attributes (saltiness, sourness, bitterness, sparkling intensity and bubbles size) using “Just About Right” scales (“not enough”, “just right”, “too much”). In this way, the overall liking scores were not biased by the normative descriptive questions.

1.1.4. Meal-CLT procedure

Participants were pre-recruited and invited to attend two different meal sessions with a one-week interval between the two sessions. Each subject was sequentially presented with the two sparkling water samples in separate meal sessions (multi-session monadic sequential presentation). The sample presentation order was balanced over the panel. The tasting sessions occurred either at lunchtime between 11:00 a.m. and 2:00 p.m. (with 5 consecutive services) or at dinnertime between 6:30 p.m. and 8:45 p.m. (with 4 consecutive services). A maximum of 10 persons could attend a given session. Consumers sat alone at their table to avoid social interaction which in this case would have been unrealistic, because participants did not know each other. For both meal sessions, a complete meal with a tabbouleh, a vegetable tart with salad and a fresh fruit salad was proposed to the participants. This meal was described during recruitment and only the participants who stated that they would probably consume such a meal were selected. All meal items were freshly cooked, packaged in individual portions and delivered by a catering firm in the morning (for the midday sessions) and in the afternoon (for the dinner sessions). In the experimental kitchen, all items were prepared for presentation on plain dishes. Thus, participants received the different meal items sequentially and they were instructed to eat as much as they wished, but to taste at least all the components of the meal including the sparkling water put on their table. The sparkling water samples (50 cl) were presented in an unopened original packaging bottle without any label or identification sign. The two original packaging bottles were rather similar and noticeable differences were considered to be negligible. All people in the same group session tasted the same water product. They were told they could consume the water as they wished i.e. directly from the bottle or in a glass. A questionnaire was presented and self-completed after the meal with the instruction to give a rating between 1 and 10 (1=dislike very much; 10=like very much) describing their overall liking for each meal component (with the sparkling water asked first) as well as for their overall liking for the entire meal. A second questionnaire was then presented to collect additional information using the same “just about right” questions as used in the traditional CLT. Attribute related questions were asked for all

meal items to avoid focussing the participants' attention specifically on the sparkling water. However, in this paper we will only consider the responses given to the sparkling water items. The participants came back one week later at the same time to attend the second meal event. They then were told that all meal components had been slightly modified in comparison with the first session. In reality however, the same meal was presented, but now accompanied by the second sparkling water sample. Again, they responded to the questionnaires on overall liking and the diagnostic "just right" questions. A 50 € compensation was then given to the participants at the end of this second session.

1.1.5. Data analysis

Data analysis was performed with Jump 5.0 (the SAS Institute Inc). For all analyses, the significance level was set at 5 %. For each protocol, mean overall liking and standard deviations were calculated. The difference between the mean scores of the two products was tested by ANOVA ($score = subject + product + order$). Eventually, we used an ANOVA ($score = product + method + product \times method$) to determine the influence of the protocol on the overall liking. With regard to the diagnostic questions (JAR), the percentage of subjects who responded in each category ("not enough", "just right", "too much") was calculated and the distribution of responses in the two samples was compared using a χ^2 test.

1.2. RESULTS

1.2.1. Overall liking: traditional CLT vs. meal CLT

Table 2 summarises the mean overall hedonic ratings for each sparkling water sample obtained with the traditional CLT and with the meal-CLT. As can be seen, the outcome of the test depends on the method used. With the traditional CLT, product A is liked significantly more than product B, whereas no significant difference in liking is found with the meal-CLT.

With regard to the impact of the tasting conditions on general liking, it can be seen that the mean hedonic score obtained for both samples of sparkling water in the meal-CLT is higher than that obtained in the traditional CLT [ANOVA *method* effect: $F_{(1,638)} = 73.09$; $p < 0.001$]. The *product* \times *method* interaction is not significant [$F_{(1,638)} = 1.45$; $p = 0.228$] even though the two methods show different levels of discrimination. This non-significant interaction illustrates the fact that the impact of the tasting conditions on the difference of liking is quite low. Finally, an unexpected order effect (over-scoring of the first tasted product) was found in the meal-CLT conditions, but not in the traditional CLT.

Table 2. Mean overall liking scores and ANOVA results for the traditional CLT and the meal-CLT

	TRADITIONAL CLT		MEAL-CLT	
	Product A	Product B	Product A	Product B
Panel size	161		160	
Mean score (SD)	6.74 (1.77)	6.20 (1.98)	7.78 (1.56)	7.58 (1.80)
Subject effect	$F_{(160,159)}=1.21$; $p=0.110$ NS		$F_{(159,158)}=1.68$; $p<0.001$ **	
Product effect	$F_{(1,159)}=7.38$; $p<0.001$ **		$F_{(1,158)}=1.57$; $p=0.212$ NS	
Order effect	$F_{(1,159)}=0.96$; $p=0.330$ NS		$F_{(1,158)}=19.26$; $p<0.001$ **	

Anova results (score = subject + product + order): NS: not significant; *: p-value < 5 %; **: p-value < 1 %.

Table 3. Diagnostic questions results: repartition of responses for each product obtained in the traditional CLT and in the meal-CLT (χ^2 tests results).

		Product A			Product B			A vs. B	
		Not enough %	Just Right %	Too much %	Not enough %	Just Right %	Too much %	p-value (χ^2 test)	
Sourness	Traditional CLT	9	82	9	7	73	20	0.011	*
	Meal-CLT	16	76	8	17	65	18	0.025	*
Sparkling intensity	Traditional CLT	17	73	10	11	66	23	0.004	**
	Meal-CLT	19	71	10	13	62	25	0.003	**
Bubble size	Traditional CLT	8	79	13	9	69	22	0.080	MS
	Meal-CLT	24	72	6	14	65	21	< 0.001	**
Saltiness	Traditional CLT	4	71	25	6	72	22	0.644	NS
	Meal-CLT	14	70	16	13	71	16	0.884	NS
Bitterness	Traditional CLT	2	88	10	3	79	18	1.000	NS
	Meal-CLT	22	70	8	19	69	12	0.482	NS

χ^2 tests: NS: not significant; MS: marginally significant (p-value < 10%); *: p-value < 5 %; **: p-value < 1%

1.2.2. Diagnostic questions: traditional CLT vs. meal-CLT

Table 3 presents the distribution of responses to the samples obtained with each method. As can be seen, even if the perception of the appropriateness of the various attributes levels slightly differs from one method to another, overall, the discrimination of the two samples on these characteristics is equivalent. There are more participants who perceived product B as “too intense” in terms of sourness, sparkling and bubble size than participants who perceived these attributes as “too intense” for product A. On the contrary, the two samples were not judged differently in terms of the hedonic perception of saltiness and bitterness.

1.3. DISCUSSION

The meal-CLT situation differs in many ways from the traditional CLT situation and a number of these differences might affect the reported hedonic scores. Table 4 highlights the contextual factors that clearly differ between the two tasting conditions and that therefore may potentially affect the results when switching from the traditional CLT to the meal-CLT. Let us specify how these different tasting conditions have influenced the stated hedonic scores and responses to the diagnostic questions.

1.3.1. Overall liking: traditional CLT vs. meal-CLT

The lower scoring of the products in artificial situation tests as compared to tests carried out in natural settings has already been found in previous studies (Hersleth et al., 2003; King et al., 2004; Kozłowska et al., 2003; Meiselman, Hirsch & Popper, 1988; Murphy, Clark & Berglund, 1958). In the present experiment, the comparison of the hedonic results obtained in the two tasting conditions was carried out with two different panels. Thus, the possibility that the method effect is confounded with a panel effect can not be completely excluded. Nevertheless, it is assumed that the selection of two representative consumers samples for both conditions allows us to neglect this panel effect and to suggest several explanations which could account for this method effect (i.e. the higher scores obtained in the meal-CLT).

The consumption conditions in the meal-CLT were expected to be more appropriate than traditional CLT conditions. In fact, King et al. (2004) already showed such individual positive influence of the meal context on liking scores for beverages. During the meal-CLT, the time of the day and hence the physiological state of the participants were assumed to be appropriate to the consumption of sparkling water, which was not systematically the case during the traditional CLT.

Table 4. Status of the contextual factors that could explain different results between the traditional CLTs and the situational CLTs.

		Traditional CLT	Meal CLT	Traditional CLT	Breakfast CLT	Traditional CLT	Social CLT
Food related factors	<i>Serving size</i>	Small (18cl)	Large (50cl)	2 units (possibility to ask more)		Small	Large
	<i>Product preparation</i>	Ambient temperature		No individual preparation	Spread items	No individual preparation	
Experimental related factors	<i>Sample presentation procedure</i>	Monadic sequential mono-session	Monadic sequential multi-session	Monadic sequential mono-session		Monadic sequential mono-	Simultaneous presentation
	<i>Time of the day</i>	Throughout the day	Lunchtime, dinnertime	Throughout the day	8:00 a.m. – 11:00 a.m.	Throughout the day	8:30 p.m. – 10:30 p.m.
	<i>Social environment</i>	No social interaction		No social interaction		No social interaction	Friendly social interaction
	<i>Food environment</i>	None	Meal items	None	Hot/cold beverages	None	Alcohol and non-alcohol beverages
	<i>Dinnerware</i>	Plastic glass	Real dinnerware	Plastic plate	Real dinnerware	Plastic dinnerware	
	<i>Ambience</i>	Usual CLT ambience		Usual CLT ambience		Usual CLT ambience	Music / filtered light / decorated
	<i>Freedom of choice</i>	No choice		No choice		No choice	Choice
	<i>Effort to obtain food</i>	No effort		No effort		No effort	Effort
	<i>Reward</i>	No reward	Reward (50 Euros)	No reward		No reward	

Furthermore, the meal environment may have been especially favourable to the consumption of water which may have led to an enhancement of the satisfaction induced by drinking it. In fact, the positive digestive function provoked by the sparkling waters could have probably been felt during the meal-CLT, whereas it could have not during the traditional CLT. The cosy environment of the meal situation as well as the reward and the free meals could also have enhanced the overall well-being of the subjects and consequently their ratings (Bell, 1993). Finally, subjects probably wanted to please the test administrator for paying them to eat and gave high scores to the samples as a sign of gratitude.

Unexpectedly, the *presentation order* effect was significant for the meal-CLT (overscoring of the product tasted first) where the two evaluations were separated by one week (multi-session monadic sequential presentation) whereas it was not significant for the traditional CLT conducted in a mono-session monadic sequential way. This phenomenon may reflect the important role of memory on the hedonic impression. However, we think that the diagnostic questions may have also played a role in this *order* effect. Indeed, the two hedonic scores reported in the traditional CLT were not biased by the diagnostic questions, which were only asked at the end of the session whereas, during the multi-session meal-CLT, the hedonic scores of the second session may have been influenced by the diagnostic questions asked in the first session which probably incited the participants to be more critical with respect to the sample tasted second.

Notably, the evaluation conditions not only influenced the level of the hedonic scores, but also the conclusion inferred from the tests, which is more worrying. The traditional CLT results reveal a significant difference in acceptability between the two types of sparkling waters that the meal-CLT results does not reveal. Thus, it can be concluded that the drinking situation influences the results for the two types of sparkling waters and that this difference is due to the fact that the sparkling waters were presented either alone or within a complete meal (food environment). The traditional CLT conditions without food environment may have magnified a difference in liking that probably did not exist when the waters were consumed during the meal. This lower discrimination of the two waters under the meal condition is also consistent with the one found by King et al. (2004) for the liking of two flavoured iced teas. To a lesser extent, other contextual factors, such as rewarding of participants or sample presentation (mono vs. multi-session), may also have played a role in the lower discrimination power of the meal-CLT in comparison with the traditional CLT. For example, Bell (1993) found that differences between four samples were greater among unpaid subjects than among subjects who were paid for their participation. Furthermore, the diagnostic questions asked

between the two overall liking questions may also explain the difference in discrimination between the two samples. It has been observed that introspection about specific characteristics leads to moderated relative evaluations (Gross, 1964; Wilson & Schooler, 1991).

1.3.2. Diagnostic questions: traditional CLT vs. meal CLT

During the traditional CLT, the “too intense” sensory characteristics of product B may explain that it was less liked than product A. However, we expected that this “too intense” aspect of brand B would be less perceived while drinking the water in combination with food. In fact, during the meal, participants still perceived product B as “too intense” in terms of sourness, sparkling level and bubble size. This point reveals that, even during a meal, participants still differentiated the two water samples. Consequently, we could hypothesise that the non-significant difference in acceptability obtained in meal-CLT does not depend upon the fact that the consumers had difficulties to differentiate the products. However, it is striking that, during the meal-CLT, the negative feeling toward the “too intense” product B did not induce a difference of overall liking between the two brands. One might assume that, during the meal-CLT, the trigeminal sensations (sparkling level and bubble size) were not decisive in the overall hedonic impression formation. During the traditional CLT, participants focused their overall hedonic evaluation on the salient differences, whereas participants did not take into account these differences in their judgement when the water was drunk during a meal. Maybe in a meal environment, the good levels of saltiness and bitterness of the water samples (taste intensity) played a major role in the overall liking. Given the fact that the performances of the two products were identical on these criteria, the two products were finally equally well liked.

To conclude, it seems that participants did not focus their overall judgement on the same sensory criteria when they taste the sparkling water samples alone or within a meal-context. This information is crucial, given the fact that in France, sparkling waters are consumed during a meal in 50% of the cases (Usage and Attitude survey carried out by marketing research services of Danone).

2. EXPERIMENT 2: BREAKFAST CONTEXT

2.1. MATERIAL AND METHODS

2.1.1. Subjects

Participants (38 men and 87 women, ages 18-65) were recruited among students (63%) and staff members (37%) of ENSIA (AgroParisTech). Sixty-one subjects (72% of females) participated in the traditional CLT experiment and sixty-four subjects (67% of females) participated in the breakfast-CLT. All participants were consumers of crispy bread.

2.1.2. Samples

Two brands of crispy bread (brand A vs. brand B) were evaluated. The sensory profiles carried out on the two types of crispy breads revealed differences on a number of attributes: visual (round shape for A and oval shape for B), texture (brand A is less hard) and taste (brand A is less sweet and more salty). No information was given to the subjects regarding the products (brand, price...). Again, the samples evaluated in both conditions belonged to the same homogeneous batches.

2.1.3. Traditional CLT procedure

Throughout the day (between 10:00 a.m. and 6:00 p.m.) participants were pre-recruited to participate in a tasting session. Evaluations took place at separate tables. Each subject was presented sequentially with the two crispy breads. Order was balanced over subjects. Crispy bread samples were presented on white disposal plates. Subjects were instructed to eat as much as they wished of each sample but at least half bread of each sample. Respondents drank some tap water before evaluating the products. Then subjects were asked to rate their overall liking for the tasted crispy bread on a 10-point scale (1 = dislike very much; 10 = like very much). After having evaluated both products on overall liking, respondents sequentially received once more the two products to rate them on two partial appreciation questions (taste and shape) with the same 10-point scale.

2.1.4. Breakfast-CLT procedure

Subjects were pre-recruited and invited to attend a breakfast session between 8:00 a.m. and 11:00 a.m. They were asked not to eat before the session. They were told that two different samples would be evaluated in the session. The sample presentation order was

balanced over the panel. When they arrived, participants were asked to take a tray and to help themselves to breakfast items presented on a buffet. They were instructed to compose their breakfast as if they were at home and thus to choose their usual food items. Food items proposed were selected on the basis of usage and attitude questionnaires filled out beforehand. On these questionnaires participants had ticked the usual components of their breakfast. Thus, coffee, hot milk, cold milk, chocolate, tea and orange juice were at their disposal. Hot beverages were prepared before the session and were kept hot in thermo-regulator bottles. Various individual portion sizes of spread products were also offered: butter (salted or not), strawberry jam, apricot jam, honey and chocolate spread. After having selected the appropriate components, they sat at individual tables without social interaction, since in France the majority of breakfast events occur without social interaction. Experimenters then presented two units of the first sample of crispy bread on a plate. In order to reduce variability due to bread preparation, participants were instructed to prepare the bread as they usually did, but to stick to the same manner for all crispy breads consumed throughout the session. Thus, both samples were evaluated under the same preparation conditions: type of beverage, type of spread products and soaking (a typical French habit) or not. Participants were asked to eat as much as they wished and had the possibility to ask the experimenter for another unit of crispy bread. A self-administered questionnaire was presented with the sample to be evaluated. Participants were asked to score their overall liking for the tasted bread (1 = dislike very much; 10 = like very much). The second crispy bread sample was then presented with the same instructions and the same questionnaire. After the overall evaluations, participants were asked the same partial appreciation questions on taste and shape. While doing this, they had the possibility to taste the samples again, but under the same preparation conditions as before. Finally at the end of the session, participants had to describe the preparation conditions they had chosen to consume the breads in a questionnaire.

2.1.5. Data analysis

For both protocols, mean scores and standard deviations were calculated (overall liking and partial appreciation). The differences between the mean scores of the two products (concerning overall liking and partial appreciation) were tested by ANOVA (*score = subject + product + order*). Eventually, we used an ANOVA (*score = product + method + product × method*) to determine the influence of the protocol on the overall liking. Considering that these data were collected on the basis of small panel sizes, it was decided to set the significance level at 10% for all analyses.

2.2. RESULTS

2.2.1. Traditional CLT vs. breakfast-CLT

Table 5 summarises the results of the ANOVAs carried out on the liking scores for both the traditional CLT and the breakfast-CLT.

Table 5. Mean liking scores (overall and partial) and ANOVAs results for the traditional CLT and the breakfast CLT

	TRADITIONAL CLT		BREAKFAST-CLT	
	Brand A	Brand B	Brand A	Brand B
Panel size	61		64	
Overall liking				
Mean score (SD)	6.81 (1.52)	7.23 (1.63)	6.66 (1.62)	7.01 (1.46)
Subject effect	$F_{(60,59)}=1.38; p=0.106$ NS		$F_{(63,62)}=2.45; p<0.001$ **	
Product effect	$F_{(1,59)}=2.62; p=0.111$ NS		$F_{(1,62)}=2.47; p=0.121$ NS	
Order effect	$F_{(1,59)}=3.96; p=0.051$ *		$F_{(1,62)}=1.26; p=0.265$ NS	
Taste liking				
Mean score (SD)	6.33 (1.88)	7.31 (1.41)	6.61 (1.86)	6.56 (1.96)
Subject effect	$F_{(60,59)}=1.17; p=0.267$ NS		$F_{(63,62)}=1.15; p=0.281$ NS	
Product effect	$F_{(1,59)}=7.27; p=0.009$ **		$F_{(1,62)}=0.002; p=0.962$ NS	
Order effect	$F_{(1,59)}=6.33; p=0.015$ *		$F_{(1,62)}=0.021; p=0.887$ NS	
Shape liking				
Mean score (SD)	7.11 (2.59)	7.44 (1.94)	5.87 (1.90)	7.54 (1.65)
Subject effect	$F_{(60,59)}=1.09; p=0.369$ NS		$F_{(63,62)}=2.23; p<0.001$ **	
Product effect	$F_{(1,59)}=1.17; p=0.282$ NS		$F_{(1,62)}=39.16; p<0.001$ **	
Order effect	$F_{(1,59)}=2.02; p=0.160$ NS		$F_{(1,62)}=0.31; p=0.579$ NS	

Anova results (score = subject + product + order): NS: not significant; *: p-value < 5 %; **: p-value < 1 %.

It can be seen that, for both protocols, the mean overall hedonic score of product A is slightly lower than that of product B. However, these two differences are not significant. The impact of the tasting conditions on the overall liking is tested with ANOVA and we only found a significant *product* effect [$F_{(1,246)} = 3.47; p = 0.063$]. Contrary to previous studies comparing hedonic scores obtained either in a natural eating environment or under artificial conditions, we did not find a significant *method* effect [$F_{(1,246)} = 1.07; p = 0.303$]. Finally, the non-significant *product* × *method* interaction [$F_{(1,246)} = 0.04; p = 0.836$] shows that the differences between the mean overall liking scores are similar for the two methods.

In contrast, the partial hedonic scores (Table 5) lead to more insightful results. It can be seen that, this time, the two tasting conditions lead to different results.

In the traditional CLT, the taste of product B was significantly more liked than that of product A. On the contrary, during the breakfast session, the tastes of the two samples were liked equally well. The results of the ANOVA on the taste liking scores show a significant *product × method* interaction [$F_{(1,246)} = 3.59$; $p = 0.059$].

The liking of the shape also led to different results depending on whether it was evaluated in the traditional or in the breakfast-CLT conditions [*product × method* interaction: $F_{(1,246)} = 7.99$; $p = 0.005$]. During the traditional CLT, participants did not report a difference of liking between the two types of shapes (round or oval), whereas, in the breakfast tasting conditions, the round shape was significantly less liked than the oval shape.

2.2.2. Breakfast-CLT: further analysis

The analysis of the breakfast consumption condition questionnaires made it possible to determine the way in which the participants chose to consume the crispy breads. It shows that all participants drank a beverage during the tasting session. In addition, nobody consumed the breads plain. All used a spread on their bread. Interestingly, 38% of the subjects soaked their crispy bread in the beverage. Given the high level of this soaking behaviour in the French population (50% according to Ipsos data), it seemed interesting to check whether this consumption behaviour influenced the hedonic scores results or not.

Three ANOVAs (*score = product + soaking + product × soaking*) were carried out on each type of hedonic question (overall and partial liking). We only found a significant *product × soaking* effect on the overall liking scores [$F_{(1,124)} = 3.68$; $p = 0.058$]. This significant interaction indicates that the difference in liking between the two products is different for the soaking or the non-soaking status. In fact, the ANOVA (*overall liking score = subject + product + order*) showed a significant *product* effect for the “soakers” [$F_{(1,22)} = 10.49$; $p = 0.004$] who scored the product B significantly higher than the product A. In contrast, for the “non-soakers” the *product* effect is not significant [$F_{(1,41)} = 0.06$; $p = 0.814$] indicating that the two products were equally liked by them.

2.3 DISCUSSION

The contextual factors which differed between the traditional CLT and the breakfast-CLT are reviewed in Table 4. The recorded consumption conditions during the breakfast-CLT clearly illustrated that eating conditions (100% of drinking and spreading of the breads) differed quite substantially from those in the traditional CLT (neither drinking nor use of spreads). Furthermore, the overall liking scores in the breakfast-CLT are based on a larger amount of bread consumed than those in the traditional CLT. The mean amount of bread consumed by participants during the breakfast-CLT before scoring the product is 1.35 units (SD=0.52), whereas it corresponds to 0.86 units (SD=0.44) during the traditional CLT. This is probably due to the fact that participants were more hungry in the breakfast-CLT than in the traditional CLT. However, despite these numerous tasting condition differences, overall liking scores were not very affected by the method. In contrast, Matuszewska et al. (1997) found an influence of product preparation by comparing imposed or free spreading procedure for consumer assessment of margarine samples. This study showed an important effect of product preparation on liking based discrimination: the more individual the approach, the better was the discrimination between the products.

Contrary to the meal context experiment, the product acceptance was not lower scored in the standardised situation than in the more appropriate situation. It should be noted that participants (students and professor) had to wake up earlier than usual to attend the early morning session. The free breakfast was probably not a sufficient reward to compensate this effort and the missing financial reward may have prevented from the enhancement of the hedonic scores that we observed in the meal-CLT. Similarly, the influence of the more naturalistic tasting condition of the restaurant-CLT set up by King et al. (2007) was insufficient to modify the level of hedonic rating in comparison with a traditional CLT.

In our experiment, the tasting conditions influenced neither the level of overall hedonic scores nor the difference in liking between the two products that did not become significant with either method. However, further analysis of the breakfast-CLT data shows that the soaking behaviour influenced the difference between the overall likings for the two products. In fact, the data of the “soakers” revealed a significantly higher liking for product B whereas the “non soakers” liked the two products equally well. The traditional CLT conditions and the breakfast-CLT conditions hence led to the same outcome only for the “non-soaking” participants.

Besides, the comparison of the partial liking scores shows a significant *product × method* interaction for both the taste and the shape. Under the traditional CLT conditions, liking for the taste of B was significantly higher than that for the taste of A, whereas the tastes of the two products were appreciated equally well in the breakfast-CLT conditions. When prepared with spread items and potentially soaked in a beverage, the two samples may lose some differences in taste and as a consequence in liking. The results of the shape-liking question are also informative. The traditional CLT yields a non-significant difference between the two products, whereas the breakfast-CLT conditions clearly shows a higher liking for the oval shape of product B. Participants may have interpreted the same question in two different ways. In the traditional CLT, liking for the shape was probably evaluated with regard to its visual appearance, whereas in the breakfast situation, participants probably evaluated the convenience of the shapes (for spreading and for soaking in the cup).

3. EXPERIMENT 3: SOCIAL CONTEXT

3.1. MATERIALS AND METHOD

3.1.1. Subjects

Two different groups (180 students age 18-30: 71 men and 109 women) were recruited to attend either the traditional CLT session (19 men and 41 women) or the social CLT session (52 men and 68 women). Participants of the social session were recruited on the campus. The subjects for the traditional CLT session were recruited among the students of ENSIA (AgroParisTech). All participants were consumers of salted crackers.

3.1.2. Samples

We assessed the degree of liking for two brands of salted cheese crackers: the national brand vs. a private label copy. The two types differed in terms of visual, texture and taste attributes. No information about the products (brand, price...) was given to the subjects.

3.1.3. Traditional CLT procedure

Participants were pre-recruited to participate in 20 min tasting sessions throughout the day (between 10:00 a.m. and 6:00 p.m.). The use of individual booths prevented social interaction between the participants. Each subject received the two crackers samples

sequentially in a balanced order. After drinking some tap water, subjects were instructed to eat as much as they wished of the presented sample, but at least 1 cracker unit. Then subjects were asked to rate their overall liking for the tasted cracker on a 10-point scale (1 = dislike very much; 10 = like very much).

3.1.4. Social CLT procedure

The protocol design of this last experiment was derived from the “time and frequency analysis” proposed by Köster (1981). The principle is based on behavioural observation and consists of organising parties at which the food and drinks to be tested are available at an *ad libitum* basis. Given the specificity of the eating context of the proposed cracker products, it was decided to test whether a modified time and frequency analysis would lead to the same hedonic information as hedonic ratings obtained under traditional CLT conditions. Thus, an evening party was organised. The locality of the party was open between 8:30 p.m. and 22:30 p.m. At the time they arrived, each of the subjects received a coded sticker and had to rate how hungry they were, using a five-point scale (from “I am not hungry at all” to “I am extremely hungry”). They were instructed to eat and drink any item they wanted but not to fetch items for other participants. The atmosphere at the party was very friendly with music and filtered light. Furthermore, alcohol (beer and planter’s punch) and non-alcohol beverages (soda, iced tea and fruit juices) were at their disposal. Participants knew each other and therefore social interaction was high. Concerning the food environment, it was decided to allow a real choice by proposing three types of salted snacks: the tested cheese crackers, chips and another kind of salted crackers. In the design, this choice possibility is quite important because it means that, for once, the participants who ate the tested crackers were not forced. Each of the snacks was presented in small coded containers with about 5g of product. Containers were presented on two tables arranged at two different locations of the room to allow the participants an easy access to the food. Both tables were divided into two areas: one where the national brand cracker was proposed and the other where the private label was proposed. Both areas proposed the tested crackers among the other snacks. Each time a participant got any tested item (national brand or private label sample), observers noted the participants’ code, the item he got and the corresponding time.

3.1.5. Data analysis

We analysed the hedonic scores of the traditional CLT in the same manner as the two other traditional CLTs with an ANOVA ($score = subject + product + order$).

Analysis of the social CLT data was performed both for the overall panel as well as for different groups of participants. Given the fact that participants had the choice to eat either the national brand or the private label sample, a choice-based analysis was used to explore the data. Furthermore, the protocol allowed exploration of the data on a dynamic basis.

Overall panel analysis: The total amount of crackers consumed during the session was calculated on the basis of the number of small containers gotten by the participants. The proportions of choices for each sample (national brand and private label) were compared using a Binomial test.

Subgroup analysis: Two different choice behaviours with regard to the tested samples can be observed: those who only eat one kind of cracker (the national brand cracker or the private label cracker) and those who sequentially eat both types of cheese crackers. The repartition of these two behaviours in our population was determined. For both groups of participants, the proportions of national brand cracker choice and private label cracker choice were compared using a Binomial test.

Dynamic analysis: Finally, we explored the development of the repartitions of the national brand and private label choices over the sequence of choices to investigate the difference between the first choice and the other choices.

3.2. RESULTS

3.2.1. Overall data

The difference between the mean hedonic score of the national brand (mean = 7.30; SD = 1.37) and that of the private label (mean = 7.25; SD = 1.63) obtained with the traditional CLT design was tested by ANOVA. The non-significant *product* effect [$F_{(1,58)} = 0.075$; $p = 0.823$] indicates that the two brands were equally well liked in the traditional CLT conditions.

As a result of the use of a free-choice protocol in the social CLT, 35 of the 120 participants did not eat any of the two tested items, thus leading to a sub-sample of 85 subjects who actually consumed the crackers for this test. Overall, these 85 participants showed a significant preference (Binomial test, $p = 0.008$) for the national brand (142 containers = 57.9%) over the private label (103 containers = 42.1%). Thus, the behavioural data reveal a preference for the national brand over the private label that the declarative data of the traditional CLT do not suggest.

3.2.2. Sub-group analysis of the social CLT data

Our social CLT design was somewhat different from the “time and frequency analysis” described by Köster (1981) because all participants had simultaneous access to the two tested samples. The two types of consumption behaviour described above were distributed as follows: 41 participants ate the two types of crackers (“multi-tasters”) and 44 participants only ate one type of crackers (“mono-tasters”). We investigated whether the hedonic conclusions inferred from these two groups of data are identical or not.

There are about as many “mono-taster” participants (Binomial test, $p = 0.326$) who only consumed the national brand (N=24) as “mono-taster” participants who only consumed the private label (N=20). Nevertheless, there are marginally (binomial test, $p = 0.051$) more national brand containers (59.5%) chosen by the “national brand mono-tasters” than private label containers (40.5%) chosen by the “private label mono-tasters”. The hunger state assessed at the beginning of the session was similar for the “national brand mono-tasters” and for the “private label mono-tasters” indicating that the over-consumption of the national brand sample was not due to hunger of the subjects.

The 41 “multi-tasters” chose to consume significantly (Binomial test, $p = 0.041$) more often the national brand (57.1%) than the private label (42.9%). This is similar to what is observed with the “mono-tasters”.

“Multi-tasters” consumed at least one of each sample. It is thus possible to compare the amount of each sample consumed on an individual basis. In fact, there are more participants who consumed predominantly national brand samples (N=18) than participants who consumed predominantly private label samples (N=7). However, a large number of participants ate the same amount of both samples (N=16). As a result, the distribution deviates only marginally from a distribution based on chance ($\chi^2 = 4.92$; $p < 0.10$). In fact, the participants who ate the same amount of both samples get only one container of each product. The fact that they did not choose a third sample prevents us to detect a preference toward one sample even if it exists.

3.2.3. Dynamic analysis

In the social CLT, the two crackers were presented simultaneously allowing the participants to go freely from one product to the other. This design allows the collection of dynamic hedonic information by exploring the development of the choices over the session. We carried out such an analysis with the data of the 60 participants who ate at least two

containers of cheese crackers. Figure 1 illustrates the proportion of national brand and private label choices according to choice position.

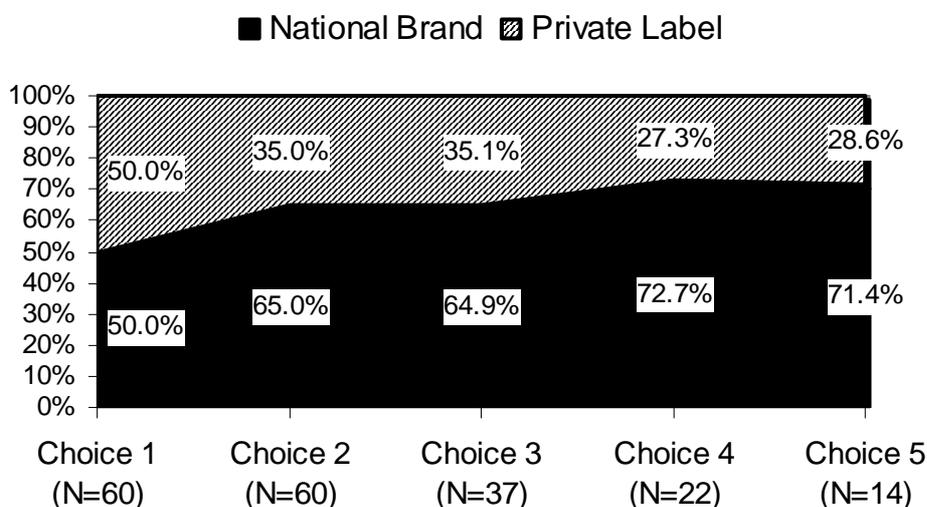


Fig. 1. Repartition of the national brand and private label choices according to the choices sequence.

As can be seen in Figure 1, in the first choice, exactly as many national brand samples as private label samples are chosen. This result gives credence to the fact that the two samples were equally accessible. The second choices show that there are more participants who chose the national brand than participants who chose the private label. The third, fourth and fifth choices confirm the formation of a clear preference for the national brand over the private label. In fact, participants who consumed more than two containers chose in majority the national brand.

3.3 DISCUSSION

3.3.1 Traditional CLT vs. social CLT

The hedonic scores collected in the traditional CLT and the intakes measured in the social CLT lead to different hedonic outputs. The social CLT revealed a significant difference of amount consumed between the two types of tested crackers whereas the traditional CLT hedonic ratings showed no significant difference of liking. Obviously, the type of hedonic measurement in the traditional CLT (hedonic scores) differs considerably from the type of hedonic measurement in the social CLT (amount consumed). Nevertheless, a great number of contextual factors also differ between the two protocols as mentioned in Table 4.

First, samples were presented in a monadic sequential way during the traditional CLT, whereas they were simultaneously presented during the social CLT. Delarue, Rason, Venturin, and Sieffermann (2001) already found that products were better discriminated when they were tasted within a simultaneous presentation than within a monadic sequential presentation.

Furthermore, the different protocol designs implied that participants ate a larger quantity of products in the social CLT than in the traditional CLT. Thus, the output of the traditional CLT ratings probably only reflects the first hedonic impression of the samples, whereas the behavioural data of the social CLT probably better reveals the long-term acceptability of the products.

Finally, during the social CLT, participants had the choice to consume the crackers or not and they had to make an effort to access them (even though it was a minimal effort). These conditions may have increased the natural impact of hedonic expectations about the effects to be experienced when eating the samples. In traditional CLT, participants certainly did build up few expectations (they had no choice and made no effort) and, not surprisingly, this led to correct hedonic impressions whatever the tasted sample. On the opposite, one may assume that more expectations were developed during the social CLT, probably leading to dissonant sensations when the product was experienced. Along this line of thinking, it is likely that the private label induced more deceptions than the national brand. Further detailed analysis of the social CLT data (see below) corroborated this latter explanation.

3.3.2. Sub-group analysis of the social CLT

The social CLT revealed two different types of consumption behaviour (mono-tasters and multi-tasters) that were separately analysed. Interestingly, both groups of participants yielded the same result, i.e. a higher consumption of the national brand than of the private label.

Several explanations may account for the strong mono-tasting behaviour. In the first place, numerous “mono-tasters” (N=25) ate only one container of crackers. Moreover, we observed thanks to the entrance questionnaire, that, overall, those who ate only one container were in fact significantly less hungry than those who ate more than one container, regardless of the eaten sample ($t = 3.22$; $p = 0.002$). The mono-tasting behaviour thus seems to be the result of a lower appetite. However, some “mono-tasters” (N=19) ate several containers of the same crackers indicating a higher appetite. These participants may have stuck to the same type of crackers, because a high level of satisfaction provoked by the first consumed cracker

led to a loyal behaviour during the whole session. In fact, there are more “national brand mono-tasters” (N=14) than “private label mono-tasters” (N=5) who ate more than one container. The overall lower consumption of the “private label mono-tasters” might thus be due to boredom caused by the recipe or to a higher tendency to provoke specific sensory satiety.

The sequential consumption of the two brands of crackers (“multi-tasters”) may reflect two different motives. First, “multi-tasters” may be people who tried the presented alternatives just out of curiosity, independently of their hedonic impression of the first tasted sample. In fact, the test conditions (coded samples, observers) could have reinforced the feeling of formal testing and thus have produced such inquisitive behaviour. Secondly, the “multi-taster” behaviour may have resulted from a real deception with the first sample, which led to a switch to the consumption of another sample. The analysis of these switches is discussed below. Due to the proximity of the two kinds of crackers on each table and since the first choice results were balanced for the two crackers, an impact of non-accessibility of one type of crackers could be safely excluded. This non-significant result of the first choice might also show that the appearance of one product was not more attractive than that of the other. The second choices are more informative than the first ones, because they may have resulted from the first organoleptic experience. The fact that the national brand was significantly more often chosen than the private label in the second choices seems to be a good indicator of the difference in overall hedonic impression induced by the two brands. It indicates that more participants who first consumed the private label decided to switch toward the national brand than vice versa. This suggests that switches were probably induced by a deception with the first chosen product rather than by mere curiosity behaviour invoked by the test conditions. We could thus conclude that the private label recipe is more deceptive than the national brand recipe.

3.3.3. General considerations

Obviously, it is difficult to know to what extent the difference in hedonic output between the social CLT and the traditional CLT was due to the tasting conditions or to the use of the two different hedonic measurements. Furthermore, in this case, it is difficult to dissociate the various contextual factors that might explain this difference in results. Contrary to the two other eating situations (meal and breakfast) where the food combination was certainly the most influential contextual factor, the social event is in fact a combination of several aspects, which has wider implications than just allowing a contact between

participants. In fact, many different contextual factors interact during a social event (e.g., comfort, long period, friendly and relaxed atmosphere). Feunekes, De Graaf, and Van Staveren (1995) and Pliner, Bell, Kinchla, and Hirsch (2003) showed no direct effect of the number of other participants on intake. They suggested that social facilitation effects on eating might be influenced more by the duration of the meal than by the presence of other people. Clendenen, Herman, and Polivy (1994) found that the relationship between meal companions is an important factor contributing to social facilitation. Eating with family or friends has more of an impact than eating with strangers. In the same way, Pliner et al. (2003) showed that a social facilitation effect occurs especially when the groups are created naturally. In view of this and as Stroebele and De Castro (2004) already pointed out, the set of contextual factors is probably closely interwoven in a social event and may be called “ambience” factors. It seems evident that the re-creation of a social atmosphere requires the manipulation of many more “ambience” factors than just allowing socialisation between participants.

4. CONCLUSION

These three experiments illustrate that tasting conditions have a significant impact on the hedonic information collected in CLT. However, this contextual influence appears in different ways depending on the experiment involved. The first experiment shows the influence of the food environment on liking for the tested beverages. Thus, if one wants to better predict the liking for this kind of product in the real world where they are usually consumed as part of a meal, one should implement consumer tests under meal conditions. It is worth noting that the tasting conditions did clearly not influence the results of the diagnostic questions. This indicates that sensory properties of the two products were finally hedonically perceived in the same way in both the traditional test and the meal-test contexts. Nevertheless, it could be assumed that, depending on the testing conditions, participants used different sensory criteria to evaluate whether the sparkling water brands were satisfying or not. In the second experiment, overall hedonic scores obtained from the traditional CLT and from the breakfast-CLT regarding the two crispy bread samples were not significantly different. Interestingly, in our experiment, the more realistic tasting conditions influenced neither the level of acceptance ratings nor the discrimination in liking of the samples. However, the analysis of the diagnostic questions data revealed that the preparation of the products (spread items) and their usage (soaking behaviour) had a strong influence on the hedonic perception

of the taste and of the shape of the products. One could question the relevance of focusing on the taste of neutral basic products themselves if one considers that very few people eat crispy bread without spread items in natural eating conditions. Besides, the traditional CLT tasting conditions induced a judgement of the shape that was based on visual perception only whereas, under real tasting conditions, it was probably based on its convenience. This latter information is important if product developers want to improve some specific convenience aspects of the products. In this specific case, one could change the shape of product A from round to oval, which is more appropriate for spreading and soaking than the round shape. Furthermore, the breakfast-CLT showed the impact of the soaking behaviour on the overall liking data. Considering that 50% of the French population soak bread in hot beverages during breakfast, one should at least take the opinion of these consumers into account when breakfast bread products are tested under realistic conditions. The last situational CLT was designed to mimic a social event based on a “time and frequency” protocol. We found that the hedonic scores obtained under traditional CLT conditions led to results that differed from those based on the amounts consumed under the social CLT conditions. Furthermore, the social CLT revealed a number of insightful results such as the development of the preference for the national brand along the eating session even though at the beginning of the session the two crackers were chosen equally often.

The present study was carried out also in order to test the feasibility of the new protocol designs under budget and logistical constraints. Seen from such an angle a number of practical conclusions can be drawn. Even if the kind of meal CLT proposed here is quite simple to manage, it is costlier than a traditional CLT, mainly because of the pre-recruitment and the rewarding of the participants. In addition, the restriction of the tasting sessions to the meal periods as well as the multi-session sample presentation procedure make it quite time-consuming in comparison with the traditional CLT. This last aspect could be improved by presenting the two beverage samples in the same session, i.e. within the same meal event. Further studies focusing on the simplification of the meal CLT will be needed before a type of meal CLT can be implemented regularly. A continental breakfast eating situation is easy to implement and only necessitates the availability of some hot beverages and some spread items. In addition, it seemed that, for this type of product, the time of the day did not play a major role in the overall impression. This suggests the possibility to alleviate the constraint of conducting the breakfast tasting sessions only in the morning. The social eating ambience was by far the most complicated eating situation to re-create. In fact, it requires a long period of

consumption in a friendly and relaxed atmosphere. Testing facilities in campuses seem to be appropriate locations for this kind of test by making it easier to gather participants who know each other and are used to attending parties. However, the exclusive recruitment of students severely reduces its representativeness. With older people, this kind of test would require an activity that prevents participants from focusing their attention on what they are eating. For example, pre- and post-event weighed individual amounts of snacks and/or beverages provided during the showing of films in old age clubs or during bus tours could be a way to address this issue. Some methodological suggestions might alleviate the design of such social event CLT and, in the same time, facilitate the data analysis. Adding a free choice component by the proposal of other non-tested food items was problematic, because it reduced the consumption of the tested crackers over the whole period to only a small amount (mean number of containers = 2.88, SD = 2.22). Furthermore, the free eating design induced a diminution of the data set size (29% of participants) which could be seen as a disadvantage of such a protocol even if it certainly improves the external validity of the data. Presentation of the tested samples only and imposed tasting of the various tested products at the beginning of the session could be designed to prevent the previously cited disadvantages.

The different attempts to mimic more realistic tasting conditions in CLT designs have succeeded in incorporating some specific contextual factors in consumer test designs as suggested by King et al. (2007). However, this reductionist approach (singling out contextual factors) presupposes determination of what is the more natural eating situation for the tested food product. Food products are rarely consumed in only one kind of eating situation, which complicates the prediction of “real life” hedonic responses based on hedonic data obtained from such mono-situational CLTs. For example, the over consumption of the national brand during the social CLT situation may purely and simply reflect that this product is more appropriate than the private label for such an aperitif situation (a friendly and relaxed atmosphere with food and beverages). In contrast, the non-significant difference of liking obtained in the traditional CLT may reflect that the two products are equally appropriated for tasting in a solo-nibbling situation.

Furthermore, as Köster (2003) pointed out, perceptual situations are not exclusively defined by objective criteria, but by the subjects’ conscious and subconscious intentions. In everyday life we live in a world of meanings rather than in one of objective facts. Even if situational CLT would be a good way to measure hedonic responses in a more meaningful context than in the traditional CLT, many of the contextual variables that we manipulate are

objective properties of the context, which in real life may have different meanings for the individual. Thus, Köster (2003) suggests the use of a situation-oriented (“deductionist”) approach based on evoked, rather than on physical situations in order not to disturb the natural perception of the situation by the subjects. The idea is based on measuring the fittingness of food products to situations as it was already proposed by Schutz (1988) in the “item by use appropriateness” method. In Köster’s approach, imaginary situations are evoked with the help of auditory and/or visual means. Consumers taste the food and only rate how well this food would be appreciated in the situations and how often they would like to eat it in such situations.

AKNOWLEDGEMENT

Authors would like to thank Gaele Lecourt, Caroline Duclos and Thomas Pilorin, students of ENSIA, for their help with the implementation of the situational CLTs.

REFERENCES

- Bell, R. (1993). Some unresolved issues of control in consumer tests: the effects of expected monetary reward and hunger. *Journal of Sensory Studies*, 8, 329-340.
- Birch, L. L., Billman, J. B., & Richards, S. S. (1984). Time of day influences food acceptability. *Appetite*, 5, 109-116.
- Bonin, D., Chambres, P., & Bernard, P. (2001). Influence du contexte sur la perception des produits alimentaires. In I. Urdapilleta, C. Ton Nu, & C. Saint-Denis (Eds.), *Traité d'évaluation sensorielle* (pp. 223-232). Paris: Dunod.
- Boutrolle, I., Delarue, J., Arranz, D., Rogeaux, M., & Köster, E. P. (2007). Central location test vs. home use test: contrasting results depending on product type. *Food Quality and Preference*, 18, 490-499.
- Cardello, A. V., & Maller, O. (1982). Acceptability of water, selected beverages and foods as a function of serving temperature. *Journal of Food Science*, 47, 1549-1552.
- Cardello, A. V., Schutz, H. G., Snow, C., & Leshner, L. L. (2000). Predictors of food acceptance, consumption and satisfaction in specific eating situations. *Food Quality and Preference*, 11, 201-216.
- Clendenen, V. I., Herman, C. P., & Polivy, J. (1994). Social facilitation of eating among friends and strangers. *Appetite*, 23, 1-13.
- De Graaf, C., Cardello, A. V., Kramer, F. M., Leshner, L. L., Meiselman, H. L., & Schutz, H. G. (2005). A comparison between liking ratings obtained under laboratory and field conditions: the role of choice. *Appetite*, 44, 15-22.
- Delarue, J., Rason, J., Venturin, C., & Sieffermann, J. M. (2001). Influence of the product presentation procedure on the outcome of hedonic tests: Is simultaneous presentation an interesting alternative to monadic sequential? In: *Proceedings of the 4th Pangborn memorial symposium: A sense odyssey*, Dijon, France, 22-26 July.
- Feunekes, G. I., De Graaf, C., & Van Staveren, W. A. (1995). Social facilitation of food intake is mediated by meal duration. *Physiology and Behavior*, 58, 551-558.
- Furst, T., Connors, M., Bisogni, C. A., Sobal, J., & Falk, L. (1996). Food choice: a conceptual model of the process. *Appetite*, 26, 247-265.
- Gross, E. J. (1964). The effect of question sequence on measures of buying interest. *Journal of Advertising Research*, 4, 40-41.
- Hersleth, M., Mevik, B. H., Naes, T., & Guinard, J. X. (2003). Effect of contextual factors on liking for wine - use of robust design methodology. *Food Quality and Preference*, 14, 615-622.
- Kanarek, R. B., & Orthen-Gambill, N. (1986). Complex interactions affecting nutrition-behavior research. *Nutrition Reviews*, 44, 172-175.

- Khan, M. A. (1981). Evaluation of food selection patterns and preferences. *CRC Critical Reviews in Food Science and Nutrition*, 15, 129-153.
- King, S. C., Meiselman, H. L., Hottenstein, A. W., Work, T. M., & Cronk, V. (2007). The effects of contextual variables on food acceptability: a confirmatory study. *Food Quality and Preference*, 18, 58-65.
- King, S. C., Weber, A. J., Meiselman, H. L., & Lv, N. (2004). The effect of meal situation, social interaction, physical environment and choice on food acceptability. *Food Quality and Preference*, 15, 645-653.
- Kozłowska, K., Jeruszka, M., Matuszewska, I., Roszkowski, W., Pikielna, N., & Brzozowska, A. (2003). Hedonic tests in different locations as predictors of apple juice consumption at home in elderly and young subjects. *Food Quality and Preference*, 14, 653-666.
- Köster, E. P. (1981). Time and frequency analysis: a new approach to the measurement of some less-well-known aspects of food preferences. In J. Solms & R. L. Hall (Eds.), *Criteria of food acceptance* (pp. 240-252). Zurich: Forster Verlag AG.
- Köster, E. P. (2003). The psychology of food choice: some often encountered fallacies. *Food Quality and Preference*, 14, 359-373.
- Kramer, F. M., Rock, K., & Engell, D. (1992). Effect of time of day and appropriateness on food intake and hedonic ratings at morning and midday. *Appetite*, 18, 1-13.
- Lawless, H., & Heymann, H. (1999). *Sensory evaluation of food. Principles and practices*. Gaithersburg, Maryland: Aspen Publishers, Inc.
- Lähteenmäki, L., & Tuorila, H. M. (1994). Liking for ice cream measured with three procedures: side-by-side, after consumption and single samples. *Journal of Sensory Studies*, 9, 455-465.
- Lähteenmäki, L., & Tuorila, H. M. (1995). Consistency of liking and appropriateness ratings and their relation to consumption in a product test of ice cream. *Appetite*, 25, 189-198.
- Matuszewska, I., Barylko-Pikielna, N., Szczecinska, A., & Radzanowska, J. (1997). Comparison of three procedures for consumer assessment of fat spreads: Short report. *Polish Journal of Food and Nutrition Science*, 6, 139-142.
- Meiselman, H. L. (1992). Methodology and theory in human eating research. *Appetite*, 19, 49-55.
- Meiselman, H. L. (1996). The contextual basis for food acceptance, food choice and food intake. In H. L. Meiselman & H. MacFie (Eds.), *Food choice acceptance and consumption* (pp. 139-263). Glasgow: Blackie Academic and Professional.
- Meiselman, H. L., Hirsch, E. S., & Popper, R. D. (1988). Sensory, hedonic and situational factors in food acceptance consumption. In D. M. H. Thomson (Ed.), *Food acceptability* (pp. 77-87). London: Elsevier.
- Murphy, E. F., Clark, B. S., & Berglund, R. M. (1958). A consumer survey versus panel testing for acceptance evaluation of Main sardines. *Food Technology*, 12, 222-226.
- Pliner, P., Bell, R., Kinchla, M., & Hirsch, E. S. (2003). Time to eat: the impact of time and social facilitation on food intake. In: *Proceedings of the 5th Pangborn sensory science symposium: A sensory revolution*, Boston, United-States, 20-24 July.
- Popper, R. D., Maller, O., & Cardello, A. V. (1989). Hedonic ratings and consumption as measures of preference for a salted entree. *Chemical Senses*, 14, 739.
- Posri, W., MacFie, H., & Henson, S. (2001). Improving the predictability of consumer liking from central location test in tea. In: *Proceedings of the 4th Pangborn memorial symposium: A sense odyssey*, Dijon, France, 22-26 July.
- Randall, E., & Sanjur, D. (1981). Food preferences - Their conceptualization and relationship to consumption. *Ecology of Food and Nutrition*, 11, 151-161.
- Rozin, P., & Tuorila, H. M. (1993). Simultaneous and temporal contextual influences on food acceptance. *Food Quality and Preference*, 4, 11-20.
- Schaefer, E. E. (1979). Manual on consumer sensory evaluation. In: *American Society for Testing and Materials*, Philadelphia: ASTM.
- Schutz, H. G. (1988). Beyond preference: appropriateness as a measure of contextual acceptance of food. In D. M. H. Thomson (Ed.), *Food Acceptability* (pp. 115-134). New York: Elsevier.
- Steenkamp, J. B. (1993). Food consumption behavior. *European Advances in Consumer Research*, 1, 401-409.
- Stroebele, N., & De Castro, J. M. (2004). Effect of ambience on food intake and food choice. *Nutrition*, 20, 821-838.
- Williams, A. A. (1983). Defining sensory quality in foods and beverages. *Chemistry and Industry*, 19, 740-745.
- Wilson, T. D. & Schooler, J. W. (1991). Thinking too much: introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, 60, 181-192.

II - UNE REPONSE HEDONIQUE MESUREE SELON DIFFERENTES SITUATIONS DE CONSOMMATION EVOQUEES : UTILISATION DE LA MEMOIRE EPISODIQUE DU PARTICIPANT

II.1 - CONTEXTE ET OBJECTIF

Nous souhaitons collecter des jugements hédoniques qui prennent en compte les situations habituelles de consommation des produits testés. Comme nous l'avons vu précédemment, la mise en place de tests en salle dans des conditions physiques plus appropriées permet aux participants d'évaluer les produits de manière plus réaliste et aboutit à des résultats différents de ceux obtenus dans les conditions classiques des tests en salle. Cependant, il reste deux problèmes non résolus :

- tous les individus ne consomment pas forcément le produit de la même façon,
- un même individu peut consommer un produit lors de différentes occasions de consommation.

Nous allons décrire ces deux problèmes dans le cas des tests en salle traditionnels et que les tests situationnels précédemment évoqués ne peuvent pas résoudre. Puis nous aborderons les solutions méthodologiques que nous proposons pour prendre en compte ces deux aspects dans la collecte des données hédoniques.

II.1.1 - Diversité des conditions de consommation et de leur signification

Il est illusoire de croire qu'une situation n'est décrite que par un ensemble de caractéristiques physiques et objectives. Les situations sont également définies par les significations attribuées à cet environnement par le biais des attentes des individus (intentionnalité). En effet, comme le précisent Drifford *et al.* (1995), l'individu doit être considéré comme un être historique doté d'une mémoire autobiographique. Cette mémoire autobiographique fait partie de la mémoire épisodique définie par Tulving (1984) comme composée d'événements ou d'épisodes personnellement vécus, localisés, datés et spécifiques à chaque individu. Tous ces événements ou épisodes sont traités par l'individu, encodés et transformés en une trace en mémoire. Ces épisodes sont alors naturellement activés lorsque l'individu est soumis à un contexte présentant des indices similaires à ceux présents dans la situation en mémoire.

Dans la perspective des tests hédoniques situationnels présentés précédemment (cf. paragraphe B-I), la situation de consommation imposée peut être éloignée de ce que l'individu

expérimente habituellement chez lui entouré de ses proches. Ceci peut donc entraîner une distorsion des signaux d'interprétation empêchant alors le sujet de récupérer en mémoire son expérience épisodique concernant la consommation de ce type de produit. Köster (2003) souligne l'importance de l'intentionnalité associée à une situation au même titre que l'environnement physique. Cette intentionnalité permet à l'individu d'allouer naturellement une signification aux objets. Köster (2003) cite Jean-Paul Sartre pour illustrer l'importance de l'intentionnalité dans nos situations quotidiennes : « *Imaginez que vous entrez dans une salle avec un tableau et un clou car vous souhaitez l'accrocher au mur. Mais vous avez oublié de prendre un marteau. Soudainement, toutes les choses qui vous entourent vous apparaissent comme pouvant potentiellement servir de marteau et vous analysez donc leur capacité à servir de marteau. Le verre montre son aspect cassable, les livres montrent la vulnérabilité de leur couverture, etc.. Finalement, vous prenez votre chaussure et vous utilisez la semelle. Quand vous avez fini et que vous regardez autour de vous, les choses ont alors perdu leur signification en tant que marteau ou non. Maintenant, les livres font référence à leur contenu ou à l'auteur que vous avez rencontré lors d'une soirée littéraire ou au lieu de vacance où vous les avez lus pour la première fois et le verre vous donne envie de boire compte tenu de la soif que vous ressentez suite à cet effort.* » Cette intentionnalité est un acte intrinsèque et inconscient dans lequel la mémoire joue un rôle important car elle dépeint l'objet avec tout ce que l'individu connaît de lui et tout ce qu'il a expérimenté avec lui. Lors des tests en salle traditionnels, cette intentionnalité est généralement complètement biaisée puisque le sujet déguste le produit avec comme seul objectif de se faire une opinion ce qui est rarement le cas dans la vie de tous les jours.

Le même environnement physique peut donc avoir différentes significations pour des individus ayant une histoire personnelle différente et il est important d'en tenir compte. Il nous a donc paru intéressant de permettre au sujet de récupérer des signaux situationnels personnels et liés à son vécu alimentaire. Dans cet objectif, nous avons testé la possibilité d'aider le sujet à activer sa mémoire épisodique avant de donner son opinion sur le produit dégusté. Cette activation de la mémoire épisodique, qui a lieu naturellement lors d'un acte de consommation, est généralement bloquée ou déformée par les conditions aseptisées des tests en salle. Forde et Delahunty (2004) ont déjà mis en place cette approche lors d'un test consommateurs sur des jus d'orange en leur faisant raconter et écrire en détail leur environnement normal de consommation de jus de fruits avant la dégustation des produits. Nous proposons donc d'utiliser cette approche en demandant au participant de décrire la situation dans laquelle il était la dernière fois qu'il a consommé le produit en question. Nous

espérons alors que la construction du jugement se fera en tenant compte des significations et intentions liées à la situation activée précédemment et sera donc plus proche de la construction de son jugement lors d'un acte de consommation naturel.

II.1.2 - Aspect multi-situationnel de la consommation des produits

La méthodologie des tests situationnels soulève également l'inconvénient de ne récupérer qu'une information hédonique dans un cadre mono-situationnel. Köster (1996) propose effectivement de ne pas considérer le consommateur comme un individu ayant des caractéristiques constantes mais plutôt comme quelqu'un qui joue différents rôles dans sa vie et qui a donc différentes intentions et attentes dans chacun de ces rôles. La figure 17 illustre le processus proposé par Gains (1989) concernant l'adéquation entre les besoins couverts par l'aliment et les besoins créés par un contexte.

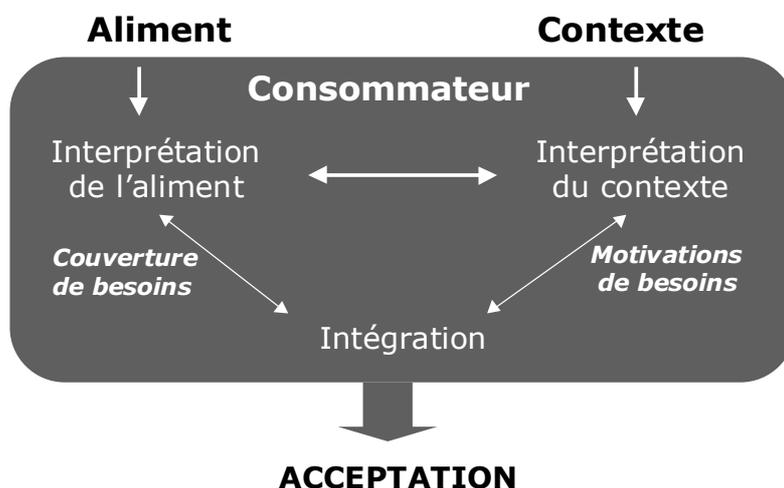


Figure 17 : Interaction entre la perception de l'aliment et la perception du contexte

La consommation d'un aliment pouvant avoir différentes intentionnalités pour un même consommateur, les études consommateurs devraient permettre de récupérer une information concernant l'appréciation des produits selon différentes situations de consommation (Miller et Ginter, 1979). Ainsi, Cardello et Schutz (1996) soulèvent l'utilité lors des tests consommateurs de récolter à la fois la façon dont un aliment est apprécié et les situations dans lesquelles il est le plus susceptible d'être consommé. Ce type de mesure hédonique récoltée en fonction de différentes situations a déjà été exploré grâce à la méthode « *item by use appropriateness* » décrite par Schutz *et al.* (1977). Il s'agit de présenter aux individus une liste d'items alimentaires ainsi qu'une liste d'usages possibles (un moment de la journée, un lieu, une occasion, un état physiologique, une personne...) et de faire noter l'à-propos de

chaque item alimentaire pour l'ensemble des usages proposés sur une échelle allant de « jamais approprié » à « toujours approprié ». Ceci permet alors de récupérer sur l'ensemble des individus interrogés une moyenne d'à-propos de chaque aliment pour chaque usage. De nombreuses études ont exploré cette méthode sur la base de listes de noms d'aliments appartenant à différentes catégories (Schutz *et al.*, 1975a, 1977) ou à la même catégorie de produits (Nantachai *et al.*, 1991 ; Raats et Shepherd, 1991 ; Schutz et Ortega, 1974 ; Schutz *et al.*, 1975b ; Scriven *et al.*, 1989 ; Sosa *et al.*, 2005). Les études ayant testé l'utilité de ce type de mesure lors de la dégustation réelle d'aliments sont plus récentes et moins nombreuses (Cardello et Schutz, 1996 ; Cardello *et al.*, 2000). Ces travaux ont souvent comparé des produits très différents entre eux ce qui rend difficile d'imaginer le potentiel d'une telle approche dans le cadre des tests consommateurs classiques qui comparent généralement des produits proches sensoriellement. Par ailleurs, cette mesure d'à-propos étant présentée comme une mesure cognitive et non affective, nous pouvons nous interroger sur sa sensibilité face à une variation de l'acceptabilité des aliments testés. Cardello et Schutz (1996) et Lähteenmäki et Tuorila (1995, 1997) ont comparé cette mesure d'à-propos avec une mesure hédonique classique et observent que des produits plus appréciés sont perçus comme globalement plus appropriés en moyenne. L'observation des données par situation met cependant en évidence des interactions entre l'appréciation des produits et leur note d'à-propos selon les situations. Bien que deux produits obtiennent des scores d'appréciation significativement différents, leurs notes d'à-propos ne vont pas systématiquement en faveur du produit le plus apprécié. Pour Lähteenmäki et Tuorila (1997), la sensibilité de la mesure d'à-propos à l'appréciation des produits est dépendante de la situation. Les notes d'à-propos de produits pour les situations les moins adaptées seraient plus stables et indépendantes de l'appréciation des produits que celles pour des situations plus adaptées. Cette stabilité du niveau d'à-propos de produits pour des situations inadéquates reflèterait une valeur culturelle et fortement cognitive liée à la catégorie de produits. En revanche, les notes d'à-propos de produits pour des situations plus adaptées présenteraient une composante affective forte et traduiraient donc le degré d'appréciation des produits. Etant donné que nous nous intéressons, dans le cadre des études consommateurs, uniquement à des situations d'usages globalement adaptées, cette sensibilité de la mesure d'à-propos à l'appréciation du produit est particulièrement intéressante et nous a donc poussé à explorer cette voie pour vérifier si, effectivement, les situations de consommation peuvent influencer la préférence vers l'une ou l'autre des recettes testées. D'après l'étude de Cardello et Schutz (1996), il est possible de récolter lors d'une seule séance des scores hédoniques et des scores d'à-propos sans que l'une ou l'autre des

mesures en soit biaisée. Cependant les travaux qui ont jusqu'à présent collecté ces données se sont généralement contentés de demander l'à-propos d'un produit pour une situation généralement décrite en quelques mots (« quand je mange devant la télévision »). Cette façon d'aborder les situations est critiquable étant donné qu'elle ne permet pas forcément au sujet de récupérer une telle situation dans sa mémoire épisodique. Par ailleurs, cela demande un effort qui focalise alors les sujets plus sur la situation que sur l'aliment dégusté risquant d'aboutir à des informations liées à la catégorie d'aliment plutôt qu'à la recette du produit. Köster (2003) souligne donc l'importance de travailler sur la façon d'évoquer les situations, la difficulté étant de permettre à tous les sujets de s'identifier à la situation évoquée et de se projeter avec leur propre expérience dans leurs situations quotidiennes.

Nous avons choisi d'explorer ce type d'approche méthodologique en permettant aux participants de construire leur opinion du produit suite à l'amorçage de plusieurs situations de consommation. Les situations de consommation étudiées doivent se forger dans la tête du sujet grâce à un mélange de mémoire et d'intentionnalité de manière à se rapprocher au mieux des signaux apportés par une situation naturelle de consommation. Des moyens suggestifs doivent donc être utilisés pour évoquer ces situations. Même si des supports visuels (images, photos, objets) ont déjà été utilisés pour évoquer des ambiances (Petit et Sieffermann, 2007), cette approche n'est pas forcément adéquate pour permettre une projection individuelle lors de l'évocation de situations globales de consommation. Nous avons fait le choix de travailler à l'aide de scénarios. Les scénarios ou « vignettes » sont définis comme de courtes histoires à propos de personnages hypothétiques dans des circonstances spécifiques ; situation dans laquelle l'interviewé est invité à répondre. Cette approche par l'intermédiaire de scénarios est depuis longtemps utilisée dans les enquêtes attitudinales notamment en sciences sociales (Finch, 1987 ; Nossanchuck, 1972 ; West, 1982) et en marketing (Bitner, 1990 ; Folkes, 1984 ; Surprenant et Solomon, 1987). Cependant, même si Carifio et Lanza suggèrent dès 1992 que les scénarios sont de bons outils pour les études qui impliquent des intentions de comportement ou des attitudes, leur utilisation commence seulement à voir le jour dans le domaine de la mesure des préférences alimentaires (Henry, communication personnelle ; Jaeger et Meiselman, 2004).

II.1.3 - Démarche générale

La méthode développée et présentée ci-après pour comparer l'appréciation de deux recettes est donc constituée de deux phases bien distinctes.

La première phase consiste en un amorçage autobiographique d'une situation de consommation individuelle au moment de la dégustation et de la construction du jugement hédonique qui permet au sujet de récupérer des signaux situationnels liés à son vécu alimentaire. Nous espérons qu'il jugera alors le produit qu'il est en train de déguster en prenant en compte les différentes variables situationnelles qui entrent en jeu naturellement dans son jugement hédonique.

Une deuxième phase est ensuite proposée de manière à prendre en compte l'aspect multi-situationnel de la consommation des produits en collectant une intention d'utilisation du produit selon différentes situations évoquées. Nous souhaitons vérifier si les produits sont appréciés de la même façon selon les situations et éventuellement identifier des situations pour lesquelles les données d'intention d'utilisation ne donnent pas les mêmes résultats que celles de l'appréciation globale.

II.2 - MATERIELS ET METHODES

II.2.1 - Les produits

Les mesures hédoniques obtenues lors du test en salle et du test à domicile réalisés sur les deux marques de crackers salés au fromage ont abouti à des résultats différents⁶ ce qui nous a poussé à remettre en question la validité externe des données obtenues en salle sur ce type de produit. Nous avons donc choisi de tester l'efficacité de cette approche dans le cadre de la comparaison de ces deux recettes A et B. La comparaison des profils sensoriels de ces deux recettes a mis en évidence de nombreuses différences en termes d'aspect, de goût et de texture. La collecte de données situationnelles nous permettra de vérifier si ces différences sensorielles conditionnent une intention d'utilisation différente selon les situations de consommation présentées.

⁶ Cf les données obtenues sur les biscuits salés présentées dans l'article 1 (Boutrolle *et al.*, 2007)

II.2.2 - Les sujets

Dans un souci de comparaison de ces résultats avec ceux du test en salle traditionnel et du test à domicile, nous avons sélectionné les sujets sur les mêmes critères que ceux utilisés pour sélectionner les participants du test en salle et du test à domicile. Ainsi, 240 femmes (âge moyen = $43 \pm 12,3$ ans), consommatrices de crackers salés au moins une fois par mois ont participé à l'étude. Elles ont été recrutées dans les rues des trois mêmes villes (Paris, Lyon et Marseille) que les consommatrices du test en salle et du test à domicile.

II.2.3 - La procédure

Une fois recrutés, les sujets sont invités à participer à une séance de dégustation en salle de 20 minutes maximum. La séance de test présente deux étapes distinctes d'évaluations auto-administrées : l'appréciation globale puis l'intention de consommation selon plusieurs situations.

Nous ne voulions pas forcer un lien individuel entre l'appréciation globale comparative des deux recettes et les intentions de consommation comparatives des deux recettes dans les différentes situations. En effet, un individu préférant largement une recette par rapport à l'autre aurait probablement conservé cet écart de notation en faveur du produit préféré pour l'ensemble des situations évoquées. Ainsi, un design de test en monadique pur a été mis en place et chaque recette a été testée par 120 sujets.

II.2.3.1 - Etape 1 : appréciation globale après amorçage autobiographique

Cette première étape consiste à projeter le sujet dans la situation dans laquelle il était la dernière fois qu'il a consommé des biscuits salés. Ce travail d'amorçage autobiographique est réalisé grâce à un questionnaire auto-administré dans lequel le sujet doit décrire sa dernière occasion de consommation. Le questionnaire a été construit de manière à ce que le sujet s'interroge sur le statut de l'ensemble des variables contextuelles mises en évidence par la littérature. La description de la situation aborde donc à la fois des variables environnementales comme le lieu, le moment, l'environnement social ou encore l'environnement alimentaire et des variables internes comme l'état de faim du sujet, son état d'humeur, etc. Pour chaque question, des propositions de réponses sont offertes ainsi que la possibilité de donner une réponse non citée dans la catégorie « autres ». La dernière question du questionnaire concerne la notation du côté agréable du moment de consommation que le sujet vient de décrire. Cette question a deux objectifs. Le premier consiste à faire prendre du recul au sujet de façon à ce qu'il ait une vision globale de la situation au moment de

construire son jugement hédonique sur le produit. Cette question a également pour objectif de vérifier l'existence ou non d'un effet de halo entre l'appréciation de la situation remémorée et les notes d'appréciation données au produit. Le remplissage de ce questionnaire se fait parallèlement à la dégustation du biscuit et une fois le questionnaire rempli, nous demandons au sujet de déguster à nouveau le biscuit et de lui donner une note d'appréciation globale sur une échelle allant de 1 à 10.

II.2.3.2 - Etape 2 : intention de consommation selon différentes situations évoquées

Choix et construction des situations

Suite à une vingtaine d'entretiens avec des consommateurs de biscuits salés, nous avons sélectionné six occasions de consommation qui correspondraient aux occasions les plus fréquentes dans lesquelles ce type de biscuits salés est consommé : apéritif en semaine, apéritif en week-end, pique-nique, grignotage pendant un moment de détente, grignotage pendant une activité de travail et grignotage dans les transports.

Les situations sont évoquées par l'intermédiaire de différents scénarios présentés dans la page suivante. La projection individuelle des participants dans les différents scénarios est réalisée à l'aide d'un support auditif. Les six scénarios ont été rédigés dans un style très simple (avec des mots simples et peu nombreux) de manière à ne pas favoriser une heuristique risquant alors d'aboutir en une interprétation des situations trop simple et liée plus à des croyances qu'à un vécu (Krosnick, 1991). Les scénarios donnent des informations sur un certain nombre de variables de contexte comme l'environnement physique mais également l'état interne du sujet. Contrairement aux études sociales qui proposent des scénarios racontés à la troisième personne du singulier, les scénarios de consommation utilisés dans cette étude sont rédigés à la première personne du singulier afin que le consommateur se projette plus facilement dans la situation décrite. Chacun des scénarios est rédigé avec un nombre de mots à peu près équivalent (78 et 87 mots) pour que la longueur du scénario n'influence pas les réponses données par les sujets.

Apéritif semaine : « Je rentre à la maison. Je suis fatigué car la journée a été longue. Je prends donc un moment pour décompresser un peu et me détendre avant le dîner. Je me prépare les composants essentiels d'un petit apéritif comme je les aime puis, je me mets à l'aise en m'installant confortablement. Je suis maintenant en train de consommer ma boisson d'apéritif préférée et je sens que j'ai une petite faim. J'attrape donc quelques biscuits salés présentés devant moi pour patienter jusqu'au dîner qui approche ».

Apéritif week-end : « C'est l'apéritif tant attendu du week-end, je peux enfin passer un peu de temps avec mes proches. Comme d'habitude, les discussions vont bon train et l'apéritif se prolonge plus que prévu. Après plusieurs verres d'apéritif, il serait peut être temps d'avaler quelque chose mais personne ne semble vouloir passer à table. Malgré le repas gargantuesque qui s'annonce, j'attrape donc un ramequin contenant des biscuits salés que je propose à mon entourage sans oublier de me servir au passage ».

Pique nique : « C'est le week-end, il fait plutôt beau. J'ai donc décidé avec mes proches de faire un pique nique à la campagne ou dans un parc près de chez moi. Après diverses activités au grand air, tout le monde est d'accord pour dire qu'il est grand temps de manger. J'étales donc au milieu de nous les ingrédients essentiels d'un bon pique nique. Après avoir réussi à récupérer les biscuits salés, je plonge ma main dans le paquet pour en sortir une pleine poignée de biscuits que je dévore ».

Grignotage détente : « Enfin seul ! Il n'y a personne pour m'empêcher de pratiquer mon activité de détente préférée comme la lecture ou la télévision. Je m'installe donc confortablement dans mon canapé muni de mon livre ou de la télécommande. Au bout d'un certain temps, la faim commence à se faire sentir. Je me traîne jusque dans la cuisine pour aller chercher un paquet de biscuits salés. De retour, je commence une longue séance de grignotage pendant laquelle les biscuits se succèdent mécaniquement dans ma bouche ».

Grignotage travail : « Tout est calme autour de moi. Je suis en train de travailler depuis déjà un bon moment. Tout ce travail de concentration, ça creuse ! Il est encore trop tôt pour faire une pause mais je ne me laisse pas abattre. Je vais bien trouver un moyen de caler mon estomac pour tenir jusqu'à la prochaine pause. Je récupère un paquet de biscuits salés qui fera bien l'affaire. Je grignote donc tranquillement quelques biscuits tout en finissant mon travail ».

Grignotage transports : « Me voilà installé dans la voiture ou le train pour un voyage qui s'annonce plutôt long et ennuyeux. Au bout de plusieurs kilomètres, la faim commence à pointer son nez. Comme d'habitude, je n'ai bien évidemment pas pris le temps de me confectionner un pique nique digne de ce nom. En revanche, j'ai eu la bonne idée d'acheter des biscuits salés ainsi que quelque chose à boire. C'est donc le moment que je choisis pour consommer tranquillement mes achats ».

Le tableau 13 récapitule l'état des variables de contexte que nous espérons communiquer par l'intermédiaire des six scénarios mis en place.

Tableau 13 : état des variables de contexte véhiculé par chacun des six scénarios

	Apéritif semaine	Apéritif week-end	Pique-nique	Grignotage détente	Grignotage travail	Grignotage transports
Lieu	domicile	domicile	extérieur, plein air	domicile	bureau ou domicile	transports
Confort	élevé	élevé	faible	élevé	moyen	moyen
Durée du moment	moyen	long	moyen	long	moyen	moyen
Moment	fin d'après midi	midi ou soir	plutôt midi	plutôt hors repas	plutôt hors repas	plutôt hors repas
Environnement social	pas forcément	oui	oui	non	non	non (inconnus)
Environnement alimentaire	oui (boisson)	oui (boisson + biscuits)	oui (boissons + aliments)	non	non	oui (boisson)
Quantité consommée	moyenne	moyenne	élevée	élevée	moyenne	élevée
Choix de l'aliment	oui	oui	oui	oui	non	non
Manipulation du produit	boite ou vaisselle	vaisselle	boite	boite	boite	boite
Etat physiologique	un peu faim, soif	faim, soif	très faim, très soif	faim	faim	faim
Etat physique	fatigué	en forme	en forme	neutre	neutre	neutre
Etat d'humeur	stressé	détendu	détendu	détendu	stressé	ennui
Evénements futurs	repas	repas	activité physique	Pas d'activité	activité cérébrale	pas de repas en perspective

Procédure expérimentale

Après avoir évalué l'appréciation globale du biscuit, chaque sujet reçoit un baladeur cassette. Les consignes suivantes lui sont alors données : « Nous allons à présent vous raconter plusieurs situations de consommation de la vie courante qui vous seront sûrement familières et auxquelles nous nous intéressons. Essayez de vraiment imaginer comment pourrait être chacune de ces situations dans votre vie quotidienne. Si vous en sentez le besoin, n'hésitez pas à fermer les yeux pour bien vous concentrer. Quand vous pensez avoir bien visualisé la première situation, faites un signe à l'enquêteur qui vous apportera un questionnaire ». Le scénario est en fait enregistré en boucle sur la cassette de manière à ce que le sujet entende plusieurs fois le même scénario. Une fois que le sujet pense avoir bien intégré la situation d'intérêt, il appelle l'enquêteur qui lui donne un questionnaire auto-administré à

remplir : « Si vous en aviez l'occasion, consommeriez-vous ce biscuit dans la situation que vous venez d'écouter ? ». Le sujet doit noter son intention de consommer le biscuit dans cette situation sur une échelle allant de 1 (« non, certainement pas ») à 10 (« oui, très certainement »). Nous avons fait le choix de demander directement au sujet son intention de consommation concernant le biscuit plutôt que de demander l'à-propos du biscuit pour la situation car nous souhaitions récupérer une information liée à la recette testée et non à la catégorie de produits. La mesure d'intention d'utilisation a déjà été utilisée en combinaison avec des contextes d'usages différents par Lähteenmäki et Tuorila (1998). Enfin, une note concernant la fréquence de la situation dans la vie du participant est demandée pour chaque situation sur une échelle allant de 1 (« jamais ») à 10 (« très régulièrement »).

Chaque sujet écoute donc les six scénarios les uns à la suite des autres (ordre de présentation équilibré) et donne son intention de consommation du biscuit ainsi que la fréquence de cette situation dans sa vie quotidienne.

II.2.4 - Analyse des données

II.2.4.1 - Etape 1 : appréciation globale après amorçage autobiographique

Les notes moyennes d'appréciation globale obtenues pour chaque recette sont comparées à l'aide d'un test de Student non apparié.

L'influence de l'appréciation du moment épisodique raconté sur l'appréciation globale des produits est testée en analysant séparément les données des sujets qui ont noté le côté plaisant de leur moment de consommation au-dessus, au niveau ou en dessous de la médiane du groupe.

II.2.4.2 - Etape 2 : intention d'utilisation selon différentes situations évoquées

Pour chaque situation, les notes moyennes d'intention de consommation pour chaque recette sont comparées par un test de Student non apparié.

Une intention globale de consommation moyenne sur l'ensemble des six situations est ensuite calculée par sujet en pondérant chaque note d'intention de consommation par la fréquence de la situation dans sa vie quotidienne :

$$\text{note pondérée individuelle} = \frac{\Sigma(\text{note} \times \text{fréquence})}{\Sigma(\text{fréquence})}$$

Ainsi, pour chaque sujet, la note d'intention de consommation du produit pour une situation peu fréquente aura moins de poids dans le calcul de la moyenne que la note d'intention de

consommation du produit pour une situation très fréquente. Ceci nous permet alors de calculer une note d'intention de consommation pondérée moyenne pour chaque produit qui sont ensuite comparées par un test de Student non apparié.

Les fréquences des situations de consommation sont également analysées par la méthode des nuées dynamiques pour mettre en évidence des groupes de sujets ayant des profils d'usage différents pour cette catégorie de produits. Les profils d'intention de consommation de chaque produit en fonction des situations sont ensuite comparés au sein de chaque groupe de sujets.

II.2.4.3 - Etape 1 vs. étape 2

La comparaison des notes d'appréciation globale et d'intention de consommation pondérée est réalisée à l'aide d'une ANOVA ($note = produit + type\ de\ note + produit \times type\ de\ note$).

Les écarts de note individuels ($\Delta_{\text{individuel}} = \text{note appréciation} - \text{note intention}$) sont calculés et comparés par groupe de participants à l'aide d'un test de Student ou d'une ANOVA selon le nombre de groupes.

II.3 - RESULTATS ET DISCUSSION

II.3.1 - Considérations méthodologiques

L'ensemble de la séance a été généralement réalisé par les participantes en moins de vingt minutes ce qui permet d'imaginer de manière relativement simple la mise en place d'un tel protocole en routine. Le questionnaire autobiographique de la première étape a été rempli en cinq minutes environ. En ce qui concerne la deuxième étape sur les situations, la phase d'écoute semble avoir tout à fait rempli son rôle. Les participantes étaient vraiment concentrées pendant l'écoute des scénarios et certaines n'ont pas hésité à fermer les yeux. Nous avons également eu des réactions verbales à l'écoute des scénarios traduisant une bonne implication des consommatrices dans cette tâche (« Ah oui, ça c'est vraiment moi ! »). Par ailleurs, les fréquences d'occurrences des situations dans la vie quotidienne des participantes permettent de valider que les situations sélectionnées correspondent à des situations de consommation quotidiennement vécues, même si les situations de grignotage dans les transports et pendant une activité de travail sont globalement moins fréquentes que les autres.

II.3.2 - Etape 1 : appréciation globale après amorçage autobiographique*II.3.2.1 - Analyse des données sur l'ensemble du panel*

Le tableau 14 présente les notes moyennes d'appréciation globale obtenues pour chaque recette en fonction de la méthode (le test en salle avec amorçage autobiographique, le test en salle traditionnel et le test à domicile) ainsi que les résultats des tests de Student associés. Les comparaisons sont réalisées sur la base des données du test en salle et du test à domicile obtenues en monadique pur (premier produit dégusté) et présentées dans l'article 1.

Tableau 14 : Scores moyens d'appréciation globale obtenus en monadique pur selon les trois méthodologies de test

	Test en salle avec amorçage autobiographique	Test en salle traditionnel	Test à domicile
Moyenne A (SD) (N=120)	7,42 (1,73)	7,51 (1,75)	8,07 (1,22)
Moyenne B (SD) (N=120)	7,43 (1,92)	7,42 (1,63)	7,47 (1,60)
t de Student (p-value)	t = 0,03 p = 0,972	t = 0,38 p = 0,702	t = 3,27 p = 0,001

Nous observons que les notes moyennes d'appréciation globale de la recette A et de la recette B collectées à la suite de l'amorçage autobiographique sont très proches et que le test de Student ne montre donc pas de différence significative d'appréciation entre les deux recettes. Ainsi, les résultats de ce test en salle avec amorçage autobiographique sont identiques aux résultats obtenus lors du test en salle classique. Forcer la récupération en mémoire d'un épisode de consommation de cette catégorie de produits n'a donc pas eu beaucoup d'effet sur l'appréciation des produits et n'a pas été suffisant pour arriver à mettre en évidence la différence d'appréciation significative qui existe probablement entre ces deux recettes et que le test à domicile a réussi à montrer. Nous pouvons alors nous demander si cette activation d'un épisode de consommation ne se produit pas de manière automatique et naturelle lors d'un test en salle traditionnel (malgré les conditions aseptisées) ce qui expliquerait les résultats très proches avec ceux du test en salle avec amorçage.

II.3.2.2 - Analyse des données par groupe de sujets

L'aspect agréable du moment de consommation récupéré en mémoire semble avoir eu un impact sur le jugement émis par les sujets concernant l'appréciation des deux recettes. La médiane des notes de l'aspect agréable du moment de consommation est de 8 (moyenne =

8,10 ; SD = 1,75). Sur la base de cette médiane nous avons donc analysé séparément les données des sujets ayant noté le moment de consommation comme plutôt moins agréable que le reste du panel (note < 8), les données des sujets ayant noté le moment de consommation comme moyennement agréable (note = 8) et les données des sujets ayant noté le moment de consommation comme plutôt plus agréable (note > 8). Le tableau 15 présente les résultats de ces trois groupes de sujet.

Tableau 15 : Scores moyens d'appréciation globale pour chaque recette selon l'amorçage d'un épisode plus ou moins agréable

		Episode moins agréable (N=75)	Episode moyennement agréable (N=56)	Episode plus agréable (N=109)
Recette A	Effectif	38	28	54
	moyenne (SD)	6,21 (1,49)	7,07 (1,38)	8,44 (1,41)
Recette B	Effectif	37	28	55
	moyenne (SD)	6,94 (1,52)	7,36 (1,81)	7,78 (2,15)
t de Student (pvalue)		t = 2,11 p = 0,038	t = 0,66 p = 0,509	t = 1,91 p = 0,060

Nous remarquons que, quelle que soit la recette, plus l'épisode de consommation raconté a été noté comme agréable, plus les notes d'appréciation globale sont élevées. Cette observation est confirmée par les résultats de l'ANOVA (*note d'appréciation = produit + groupe + produit × groupe*) avec un effet *groupe* significatif [$F_{(2,234)} = 19,12$; $p < 0,001$]. Un test de Newman-Keuls ($\alpha < 5\%$) permet effectivement de conclure que les notes d'appréciation données par les sujets ayant récupéré en mémoire un épisode plutôt agréable sont significativement plus élevées que celles données par les sujets ayant récupéré un épisode moyennement agréable qui sont à leur tour significativement plus élevées que les notes données par les sujets ayant récupéré un épisode peu agréable. Ce phénomène peut être interprété de différentes façons. Il peut traduire un effet de halo de la note de l'appréciation de l'épisode de consommation sur la note de l'appréciation du produit. Cette cohérence des réponses des sujets refléterait donc probablement de manière plus globale le plaisir de consommation alloué à la catégorie de produits testée (une occasion de consommation est peu agréable car la catégorie de produits consommée est peu appréciée). Cependant, ce phénomène peut également être tout simplement l'illustration d'un effet d'utilisation de l'échelle. Certaines participantes utiliseraient plutôt l'extrémité haute de l'échelle à 10 points quelle que soit la question (appréciation globale du produit ou de l'épisode de consommation) alors que d'autres participantes, plus exigeantes, utiliseraient plus facilement les notes plus basses de l'échelle.

Outre cette information, nous remarquons que la comparaison des notes moyennes d'appréciation des deux recettes ne donne pas les mêmes résultats selon le groupe de sujets ce qui est confirmé par l'interaction *produit* × *groupe* significative [$F_{(2,234)} = 4,11$; $p = 0,018$]. Les données des sujets ayant activé en mémoire un épisode de consommation plutôt moins agréable montrent une appréciation globale pour la recette B significativement plus élevée que pour la recette A. De leur côté, les sujets ayant activé un épisode moyennement agréable ne montrent pas de différence significative concernant l'appréciation globale des deux recettes. Enfin, de façon tout à fait surprenante, les données des participantes ayant activé un épisode de consommation plutôt agréable aboutissent également à une différence significative (à 10%) d'appréciation des deux recettes avec, cette fois, la recette A plus appréciée que la recette B. Ainsi, il semble que l'activation d'un épisode plutôt agréable de consommation de cette catégorie de produits soit en faveur de la recette A alors que l'activation d'un épisode plutôt moins agréable de consommation soit en faveur de la recette B. Cette conclusion est à prendre avec précaution étant donné que d'autres variables individuelles qui ne sont pas forcément liées à l'épisode de consommation activé peuvent également différencier ces trois groupes de sujets et être la cause de ces différences de résultats. À l'aide des réponses données au questionnaire autobiographique, nous avons tenté d'identifier des profils types de situations vécues comme plutôt agréable ou comme plutôt peu agréable par les sujets. Nous n'avons cependant pas pu dégager de descriptions communes de ces situations très agréables ou moins agréables de la part des participantes. Cette information souligne encore une fois l'importance de l'interprétation de la situation par l'individu. Contrairement aux résultats sur l'ensemble du panel, ces résultats par groupe de sujets supporteraient donc le fait que l'activation en mémoire d'un épisode de consommation a quand même eu une influence sur les notes d'appréciation globale données aux produits.

II.3.3 - Etape 2 : intention de consommation selon différentes situations évoquées

II.3.3.1 - Analyse des données sur l'ensemble du panel

Le tableau 16 présente les scores moyens d'intention de consommation (et les écarts-types) obtenus pour chaque marque de crackers selon les situations ainsi que les résultats des tests de Student associés pour chaque situation.

Tableau 16 : Scores moyens d'intention de consommation pour chaque recette selon les situations et résultats des tests de Student pour comparer les performances des deux recettes

	Apéritif week-end	Apéritif semaine	Pique-nique	Grignotage détente	Grignotage travail	Grignotage transport
Moyenne A (SD)	7,55 (2,36)	6,23 (2,89)	6,18 (2,97)	5,18 (3,31)	5,16 (3,05)	4,95 (3,26)
Moyenne B (SD)	8,13 (2,00)	6,65 (2,86)	6,30 (3,00)	5,72 (3,25)	5,16 (3,07)	5,43 (3,14)
t de Student (p-value)	t = 2,04 (p=0,043)	t = 1,14 (p=0,253)	t = 0,30 (p=0,762)	t = 1,26 (p=0,209)	t = 0,00 (p=1,00)	t = 1,15 (p=0,251)

Nous remarquons que l'intention de consommation de la recette B est supérieure à celle de la recette A dans presque toutes les situations mais que cet écart n'est significatif que dans le cas de la situation d'apéritif le week-end. L'ANOVA ($intention = produit + situation + produit \times situation$) aboutit à un effet *situation* significatif [$F_{(5,1428)} = 29,05$; $p < 0,001$] indiquant que les scores d'intention de consommation des produits sont globalement différents selon les situations. Un test de comparaison de moyennes de Newman-Keuls ($\alpha < 5\%$) montre que la performance des produits lors de la situation d'apéritif du week-end est significativement supérieure à celle obtenue dans les autres situations. Parmi les cinq situations restantes, les produits obtiennent des scores d'intention de consommation significativement plus favorables dans les situations de l'apéritif en semaine et du pique-nique que dans les trois situations de grignotage. Contrairement aux données traitées situation par situation, l'ANOVA aboutit à un effet *produit* significatif [$F_{(1,1428)} = 5,17$; $p = 0,023$] ce qui confirme que l'intention de consommation de la recette B est significativement supérieure à celle de la recette A sur l'ensemble des situations. L'interaction $produit \times situation$ non significative [$F_{(5,1428)} = 0,39$; $p = 0,857$] met en évidence que l'écart d'intention de consommation entre les deux recettes est globalement conservé quelle que soit la situation même si nous avons vu que cet écart n'est significatif que dans le cas de l'apéritif du week-end.

Dans l'objectif de récupérer une seule note d'intention de consommation globale du biscuit par individu, nous avons calculé une moyenne individuelle d'intention de consommation sur la base des six notes d'intention de consommation pondérées par les fréquences des situations dans la vie du sujet. Plus qu'une simple intention de consommation moyenne, ce calcul a l'avantage de donner, par individu, plus de poids aux intentions de consommation pour les situations très fréquentes que pour les situations peu fréquentes. Nous obtenons ensuite, sur l'ensemble du panel, une note moyenne d'intention de consommation globale pour la recette A (score d'intention moyen = 6,66 ; SD = 2,00) et pour la recette B

(score d'intention moyen = 7,16 ; SD = 1,69). Le test de Student non apparié montre ainsi une différence significative d'intention de consommation ($t = 2,10$; $p = 0,037$). Ce résultat vient donc confirmer la supériorité de l'intention de consommation générale pour la recette B par rapport à la recette A.

II.3.3.2 - Analyse des données par groupe de sujets

Les données d'intention de consommation par situation ont mis en évidence des écarts-types très élevés (de 2,00 à 3,36) illustrant la diversité des participantes en ce qui concerne l'intention de consommer les produits selon les situations. Cela nous a donc poussé à explorer les données par groupe de sujets. Pour cela, nous avons réalisé une classification des sujets grâce à la méthode des nuées dynamiques sur la base des réponses données aux fréquences d'occurrence des situations dans la vie quotidienne des participantes. Nous avons pu ainsi déterminer quatre profils différents de consommation des biscuits salés. La figure 18 représente les profils moyens des fréquences des situations des quatre groupes de sujets identifiés.

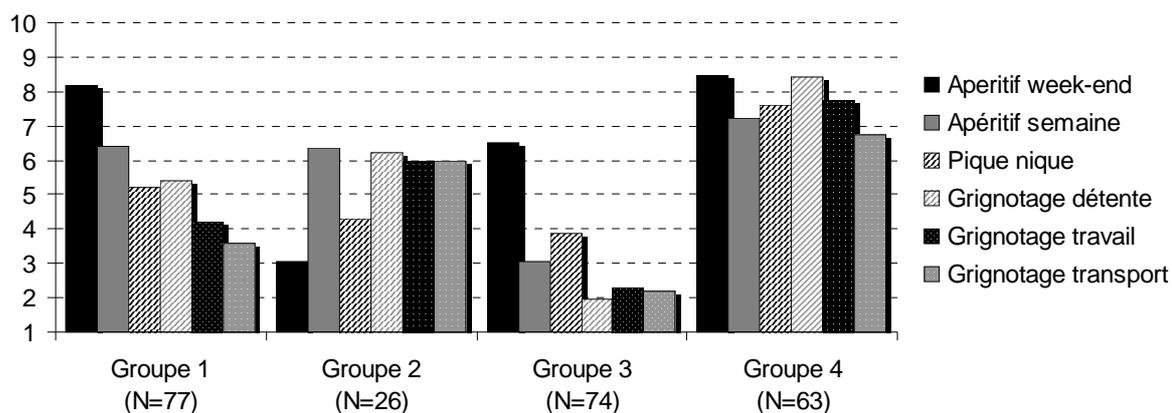


Figure 18 : Profils d'usages (moyenne des fréquences d'occurrence des situations) des 4 groupes de consommatrices obtenus par les nuées dynamiques

Le groupe 1 est constitué de sujets qui privilégient, pour ce type de produits, des situations de consommation associées à un moment social (apéritif week-end et pique-nique) ou à un moment solitaire mais de détente (apéritif semaine ou grignotage détente). Les situations pour lesquelles la consommation est plutôt utilitaire ou fonctionnelle (travail ou transport) sont moins associées à cette catégorie de produits. On parlera donc du groupe « usage plaisir ». Le groupe 2, plus restreint, représente des consommatrices qui ne consomment pas ce type de biscuits dans des environnements avec une interaction sociale forte (apéritif du week-end et pique-nique) mais uniquement lors de moments seuls (apéritif en semaine et grignotage). On regroupe donc ici des situations de consommation plutôt liées à

un moment solitaire pour lequel la consommation peut être soit strictement fonctionnelle (grignotage travail et transport) soit plutôt liée à un moment de détente (grignotage détente ou apéritif). Ces consommatrices seront identifiées comme appartenant au groupe « usage solitaire ». Le groupe 3 regroupe les sujets qui consomment ces produits principalement lors d'un apéritif le week-end (groupe « mono-usage »). A l'inverse, le groupe 4 est constitué de participantes qui consomment ces produits régulièrement dans l'ensemble des situations (groupe « multi-usage »).

L'analyse des données d'intention de consommation des deux recettes par groupe de consommatrices révèle à présent des résultats plus discriminants en ce qui concerne les deux recettes que ceux obtenus sur l'ensemble du panel.

Le groupe « usage plaisir » (figure 19) présente une intention de consommation des deux biscuits plus importante pour les situations sociales ou associées à un moment de détente. Par ailleurs, l'intention de consommation des deux recettes est significativement différente pour la situation d'apéritif du week-end et de grignotage pendant un moment de détente. Dans les deux cas, la recette B est plus performante que la recette A.

Le groupe « usage solitaire » (figure 20) présente des intentions de consommation relativement proches pour l'ensemble des situations ce qui signifie que, bien que les situations liées à des événements sociaux ne soient pas les situations qu'elles privilégient pour consommer ce type d'aliment, elles ne verraient pas d'inconvénient à les consommer dans ces situations. En ce qui concerne les situations de consommation les plus fréquentes pour ce type de produits, nous observons que, lors d'une situation de grignotage associée à un moment de détente, la recette B est significativement plus performante que la recette A alors que les recettes sont aussi performantes dans le cas d'un apéritif en semaine ou lors de grignotage au travail ou dans les transports.

Les deux autres groupes qui se différencient par le caractère mono- ou multi-usage de la consommation de ce type de biscuits présentent des profils d'intention de consommation très différents.

Le groupe « mono-usage » (figure 21) présente effectivement une intention de consommation beaucoup plus élevée en situation d'apéritif le week-end que lors des autres situations, quelle que soit la recette. Cependant, pour ce groupe de consommatrices, les deux recettes présentent globalement les mêmes scores d'intention de consommation quelle que soit la situation.

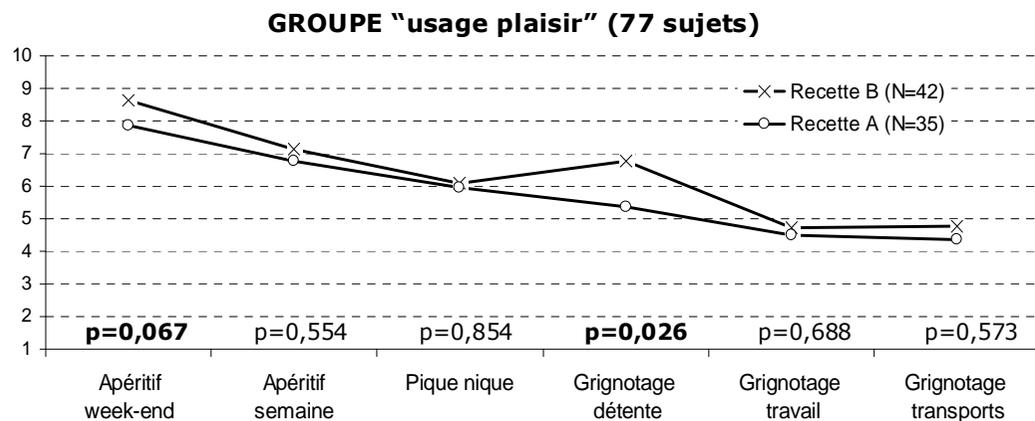


Figure 19 : Profil d'intentions de consommation pour chaque recette pour le groupe de consommatrices « usage plaisir »

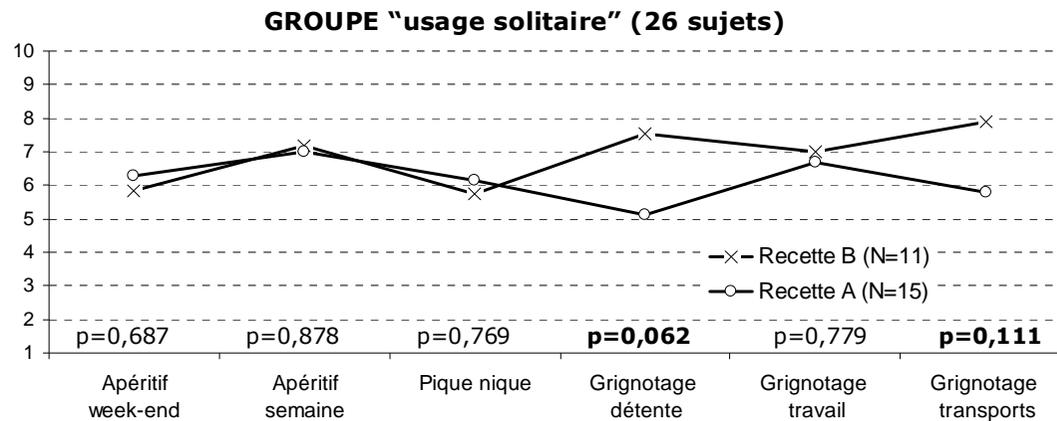


Figure 20 : Profil d'intentions de consommation pour chaque recette pour le groupe de consommatrices « usage solitaire »

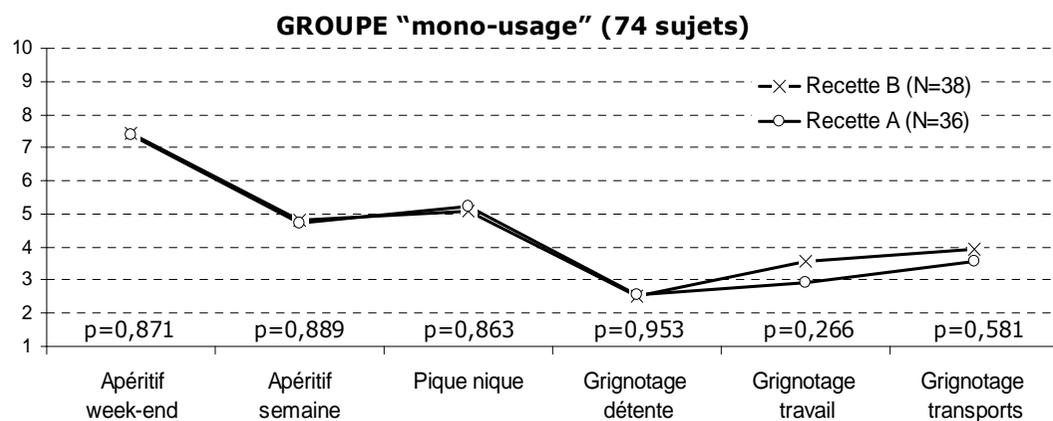


Figure 21 : Profil d'intentions de consommation pour chaque recette pour le groupe de consommatrices « mono-usage »

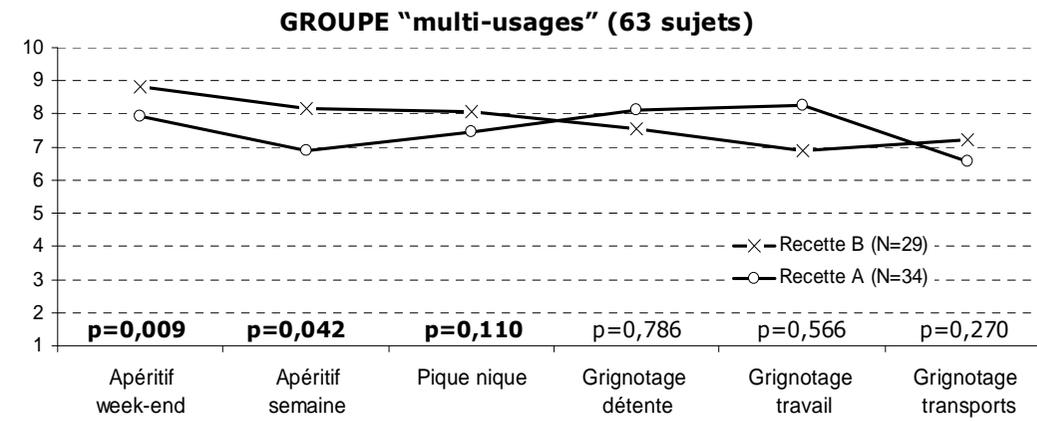


Figure 22 : Profil d'intentions de consommation pour chaque recette pour le groupe de consommatrices « multi-usage »

Les intentions de consommation des consommatrices du groupe « multi-usage » (figure 22) sont également en accord avec les fréquences des situations dans leur vie quotidienne, avec globalement des intentions de consommation élevées pour l'ensemble des situations. En revanche, nous observons une interaction en ce qui concerne la recette et les situations. L'intention de consommer la recette B est significativement plus élevée lors des deux situations d'apéritif. Les deux recettes présentent des intentions de consommation non significativement différentes pour les quatre autres situations. Cependant, nous observons que la recette B est plus performante pour le pique-nique et le grignotage dans les transports alors que la recette A est plus performante dans le cas des situations de grignotage lors d'un moment de détente ou au travail.

II.3.4 - Etape 1 vs. étape 2 : appréciation globale vs. intention de consommation

II.3.4.1 - Analyse des données sur l'ensemble du panel

Les deux étapes de ce protocole nous permettent de récolter deux types d'informations hédoniques sur chacune des recettes : la note moyenne d'appréciation globale et la note moyenne d'intention de consommation pondérée par la fréquence d'occurrence des situations. Il nous a alors paru intéressant de comparer ces deux informations hédoniques. Le tableau 17 présente les moyennes de ces deux mesures pour chaque recette.

Tableau 17 : Notes moyennes d'appréciation globale et d'intention de consommation obtenues pour chaque recette (résultats des tests de Students associés)

	Appréciation globale après activation de la mémoire épisodique	Intention de consommation globale (pondérée par les fréquences des situations)
Moyenne A (SD)	7,42 (1,73)	6,66 (2,00)
Moyenne B (SD)	7,43 (1,92)	7,16 (1,69)
t de Student (p-value)	t = 0,03 (p = 0,972)	t = 2,10 (p = 0,037)

L'ANOVA ($note = produit + type\ de\ note + produit \times type\ de\ note$) confirme qu'il existe bien des différences entre ces deux mesures.

L'effet *type de note* est très significatif [$F_{(1,476)} = 34,45 ; p < 0,001$] et les notes d'appréciation globale sont significativement plus élevées que les notes d'intention de consommation. La question d'intention de consommation basée sur des suggestions de vrais comportements (scénarios) demande aux sujets d'être peut-être plus exigeants, plus sévères

envers les produits qu'une simple question d'appréciation globale. Cette diminution du niveau de notation entre l'appréciation globale de l'étape 1 et les intentions de consommation de l'étape 2 peut également traduire un effet de lassitude étant donné que les participantes doivent re-consommer les biscuits pour répondre aux questions de l'étape 2.

La significativité de l'interaction *produit* \times *type de note* [$F_{(1,476)} = 2,08$; $p = 0,140$] s'approche du seuil de 10% traduisant des écarts de notes légèrement différents entre les deux recettes selon la mesure récoltée. Sur l'ensemble du panel, les scores d'appréciation globale des deux produits ne sont pas significativement différents alors que l'intention globale de consommation du produit B est significativement supérieure à l'intention globale de consommation du produit A. Ainsi, la mesure d'intention de consommation globale pondérée permet de différencier les produits alors que l'appréciation globale ne le permet pas.

Dans l'hypothèse d'un effet de lassitude comme explication de la diminution de la notation entre l'appréciation globale de l'étape 1 et l'intention de consommation de l'étape 2, l'interaction significative *produit* \times *type de note* pourrait également refléter le fait que la lassitude apportée par la recette A est plus forte que celle apportée par la recette B. En effet, les participantes qui ont évalué la recette A présentent un écart de note individuel moyen entre l'appréciation globale et l'intention de consommation significativement ($t = 1,91$; $p = 0,056$) plus élevé ($\Delta_{\text{moyen}} = 0,76$) que les participantes qui ont évalué la recette B ($\Delta_{\text{moyen}} = 0,26$).

II.3.4.2 - Analyse des données par groupe de sujets

Analyse par groupe de niveau d'appréciation

Lähteenmäki et Tuorila (1997) ont étudié l'influence du niveau d'appréciation de produits sur leur niveau d'à-propos en comparant les notes d'à-propos données par des sujets ayant apprécié les produits avec celles données par des sujets ayant moins apprécié les produits. Dans leur étude, le niveau d'appréciation définit le niveau global d'à-propos mais ne modifie pas énormément la forme des profils d'à-propos. Les auteurs précisent que le niveau d'à-propos est relié au niveau d'appréciation principalement pour des contextes d'utilisation culturellement adaptés. Ainsi, l'à-propos d'un produit pour une situation peut être modifié par son appréciation au-dessus d'un certain niveau mais, le fait qu'un produit soit inapproprié pour une situation semble être un concept culturellement défini qui ne peut pas être inversé par l'appréciation élevée du produit. Afin d'étudier le lien entre l'appréciation globale des recettes et les profils d'intention de consommation, nous avons procédé de la même façon que

Lähteenmäki et Tuorila (1997) en comparant les profils obtenus par les sujets qui ont noté l'appréciation globale des produits supérieure à la médiane avec ceux des sujets qui ont donné une note d'appréciation inférieure à la médiane. Dans notre cas, les situations d'utilisation suggérées ne sont jamais complètement inadéquates et nous retrouvons donc sur la figure 23 un lien fort entre l'appréciation des produits et l'intention de les consommer dans les différentes situations. Quelles que soient les situations et les produits évalués, les consommatrices qui apprécient le produit donnent des notes d'intention de consommation significativement plus élevées que les consommatrices qui apprécient moins le produit.

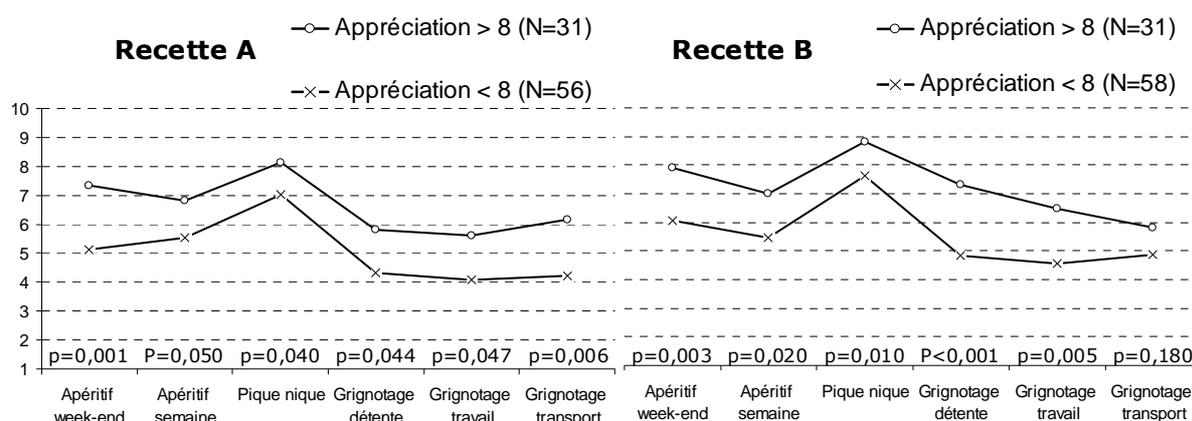


Figure 23 : Profils d'intention de consommation des deux recettes par groupe de niveau d'appréciation (*p*-values associées aux tests de Student non appariés pour chaque situation)

Outre le fait qu'un produit plus apprécié présente une intention de consommation plus élevée quelle que soit la situation, nous ne pouvons bien évidemment pas exclure le fait que cette relation puisse être également le fruit d'une utilisation différente de l'échelle de notation par les deux groupes de consommatrices qui auraient tendance à utiliser systématiquement la partie haute ou la partie basse de l'échelle quelle que soit la question posée.

Nous pouvons approfondir l'exploration du lien entre les deux types de notations en analysant, au niveau individuel, l'écart de notation entre l'appréciation globale du produit et l'intention de le consommer ($\Delta_{\text{individuel}} = \text{note d'appréciation} - \text{note d'intention de consommation pondérée}$) par groupe de participantes. Le tableau 18 présente les Δ_{moyens} des trois groupes de participantes ayant plus ou moins apprécié les produits évalués ainsi que les résultats de l'ANOVA ($\Delta_{\text{moyen}} = \text{groupe}$) et ceux du test de Newman-Keuls ($\alpha < 5\%$).

Tableau 18 : Ecart de note individuel moyen par groupe de niveaux d'appréciation des produits

	Produits peu appréciés (note hédonique < 8)	Produits moyennement appréciés (note hédonique = 8)	Produits très appréciés (note hédonique > 8)
Effectif	114	62	64
Δ_{moyen}	- 0,32 ^c	0,83 ^b	1,68 ^a
ANOVA (effet groupe)	$F_{(2,237)} = 25,37$ $p < 0,001$		

Nous observons que le lien au niveau individuel entre les deux types de notations est dépendant du niveau d'appréciation des produits évalués. Ainsi, les participantes qui ont peu apprécié les produits présentent un écart de notes plutôt faible, voire même négatif entre la note d'appréciation globale et la note d'intention de consommation. Une appréciation globale plutôt faible serait donc reliée à une intention de consommation assez proche ou même supérieure. En revanche, les participantes qui déclarent apprécier fortement les produits évalués donnent des notes d'intention de consommation beaucoup plus faibles. Dans l'hypothèse d'une diminution de la notation due à un effet de lassitude, nous pouvons supposer que les participantes qui apprécient fortement les produits lors de la première consommation sont plus fortement sensibles à cet effet que les participantes qui apprécient moins les produits.

Analyse par groupes d'usage

Nous pouvons également nous intéresser à l'impact des quatre profils de consommation précédemment identifiés sur l'appréciation globale des deux recettes. La figure 24 représente la moyenne des notes d'appréciation globale pour chaque recette selon le groupe de participantes.

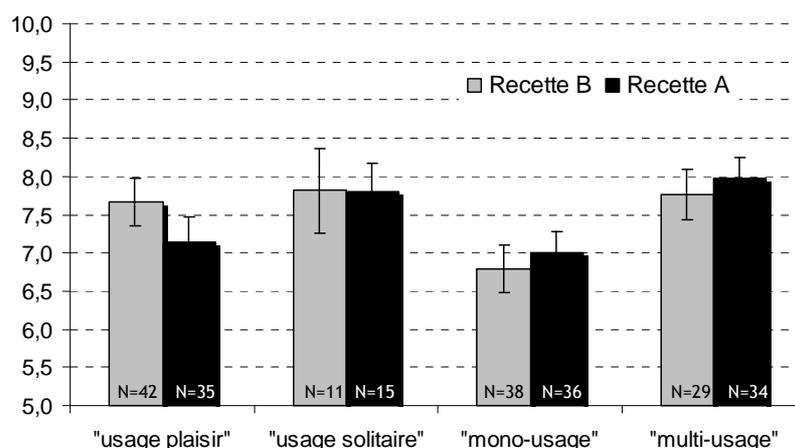


Figure 24 : Scores moyens d'appréciation globale pour chaque recette obtenus selon les groupes de consommatrices identifiés par les fréquences d'occurrence des situations

Nous observons que les quatre groupes de consommatrices notent les produits différemment. L'ANOVA (*appréciation globale = produit + groupe + produit × groupe*) met en évidence un effet *groupe* significatif [$F_{(3, 232)} = 3,87$; $p = 0,010$]. Le test de comparaison multiple de Newman-Keuls ($\alpha < 5\%$) montre que les notes d'appréciation globales données par le groupe « mono-usage » sont significativement moins élevées que les notes d'appréciation globale données par le groupe « multi-usage ». Ainsi, les consommatrices qui privilégient uniquement la situation d'apéritif le week-end pour cette catégorie de produits apprécient moins les deux produits testés que les consommatrices qui pourraient consommer cette catégorie de produits dans l'ensemble des situations. L'interaction *produit × groupe* est non significative [$F_{(3, 232)} = 0,68$; $p = 0,562$] même si l'on observe des différences d'écart de note entre les deux recettes selon les groupes. Aucune des différences d'appréciation entre les deux recettes n'est significative. Cependant, nous pouvons voir que le groupe « usage plaisir » a tendance à être plutôt en faveur de la recette B alors que les deux groupes « mono et multi-usage » auraient tendance à être plutôt en faveur de la recette A. Le groupe « usage solitaire » quant à lui ne semble pas présenter de préférence envers une recette.

II.4 - DISCUSSION GENERALE

Pour en revenir à notre comparaison de méthodes sur la base du test en salle traditionnel et du test à domicile, les résultats de ce test en salle avec amorçage autobiographique et suggestion de situations d'usages viennent encore une fois confirmer l'importance du protocole de test et du type de question dans la mise en place des études consommateurs.

La question d'appréciation globale posée dans cette étude aboutit à une conclusion différente de celle posée à domicile. Ainsi, l'amorçage autobiographique ne serait pas suffisant pour faire intervenir, lors de la construction du jugement hédonique en salle, les variables contextuelles qui interviennent naturellement lors de la construction du jugement hédonique pendant un test à domicile. Le moyen que nous avons utilisé pour activer la mémoire épisodique peut également être remis en cause. En effet, l'utilisation d'un questionnaire a été choisie dans le but de simplifier la tâche de remémoration mais son aspect trop directif a pu tout à fait perturber cette phase. Il serait peut-être intéressant de réaliser le même type de tâche de remémoration avant l'évaluation d'un produit en laissant le sujet raconter librement son épisode de consommation. L'analyse des données selon l'activation d'un moment plus ou moins plaisant semble cependant illustrer le fait que l'activation de la mémoire épisodique a quand même eu un impact sur l'appréciation des recettes avec les

situations vécues comme plaisantes plutôt favorables à la recette A et les situations vécues comme moins plaisantes plutôt favorables à la recette B. Ce résultat pourrait suggérer que les notes d'appréciation en faveur de la recette A lors du test à domicile proviennent du fait qu'il est vécu comme un épisode de consommation plutôt agréable par les participantes.

Plus surprenant encore, la question d'intention de consommation en fonction des situations met en évidence une nouvelle information hédonique concernant la comparaison de ces deux recettes : les sujets seraient globalement plus favorables à une intention de consommation concernant la recette B que la recette A. Une des premières explications de ce résultat inattendu concerne la probable perturbation de l'étape 2 par l'étape 1. En effet, la phase de remémoration d'un épisode de consommation par un questionnaire à remplir est particulièrement précise et demande une réflexion importante qui a pu amener le sujet à être dans un état d'analyse trop élevé pour la deuxième phase et différent de l'état dans lequel il peut être lors d'un test à domicile. Bien évidemment, la sélection des situations de consommation peut également avoir mis de côté une situation dans laquelle l'intention de consommation de la recette A aurait été supérieure à celle de la recette B comme dans le cas du test à domicile. Quoi qu'il en soit, le résultat contradictoire de l'intention de consommation de ce protocole de test en salle avec celui de l'appréciation globale du test à domicile soulève encore une fois le problème de la détermination de la valeur de référence. Effectivement, nous avons de manière intuitive associé les données du test à domicile aux données hédoniques de référence mais nous ne pouvons pas exclure le fait que le test à domicile puisse ne pas être le bon révélateur de l'appréciation des produits en situation naturelle de consommation. Peut-être alors que l'information hédonique sur l'intention de consommer les biscuits collectée en salle avec l'amorçage des situations serait plus prédictive de la réalité.

Par ailleurs, au sein de la même étude, nous observons également que les deux types de mesures hédoniques obtenues avec les mêmes sujets (appréciation globale et intention de consommation) n'aboutissent pas au même résultat. Ces deux mesures ne s'intéressent bien évidemment pas fondamentalement au même concept ce qui peut expliquer le fait que l'appréciation globale ne montre pas de différence significative entre les deux produits alors que l'intention de consommation moyenne pondérée par les fréquences d'occurrence des situations montre une meilleure performance de la recette B par rapport à la recette A. En fait, c'est principalement lors de la situation d'un apéritif pendant le week-end que la recette B est notée comme clairement plus favorable que la recette A. Cette situation présente une composante d'environnement social très forte qui pourrait peut-être expliquer la préférence de proposer à ses invités un biscuit plus généreux dans un souci de désirabilité sociale. En effet,

les profils sensoriels de ces deux recettes montrent que le biscuit B est plus épais, plus cuit et plus intense en goût que le biscuit A ce qui lui confère peut être un aspect plus traditionnel. Ainsi, alors que la question d'appréciation globale s'intéresse peut-être plus à l'opinion personnelle des consommatrices, la question d'intention de consommation dans les situations fait probablement intervenir une composante de représentation sociale. Lemoine (1999) a déjà mis en évidence que l'anticipation de l'environnement social au moment de la consommation du produit influence les sources d'informations externes utilisées préalablement à un acte d'achat alimentaire. Nous pouvons supposer que cette modification des sources d'information impacte également le jugement du produit. Il convient alors de s'interroger sur l'information hédonique la plus pertinente que nous souhaitons collecter lors des tests consommateurs.

La dernière information apportée par cette étude concerne la richesse des informations obtenues par groupe de sujets. Ainsi, les quatre profils d'usage concernant ce type de produits aboutissent à des profils différents d'intention de consommation selon les situations. Par ailleurs, les profils d'usage de la catégorie de produits auraient tendance à avoir un impact sur l'appréciation globale des produits mettant ainsi en évidence des groupes plutôt en faveur de la recette A (« mono et multi-usage »), un groupe plutôt en faveur de la recette B (« usage plaisir ») et un groupe plutôt indifférent à la recette (« usage solitaire »). Nous souhaitons donc insister sur l'importance de catégoriser les sujets sur la base de variables autres que le sexe ou l'âge. Les habitudes contextuelles de consommation concernant la catégorie de produits testée semblent constituer une variable particulièrement pertinente. Cette étude ne permet malheureusement pas d'analyser de manière très pertinente ces groupes de sujets qui présentent des effectifs inférieurs à 60 mais elle illustre le potentiel de telles données qui seraient obtenues sur des effectifs plus importants. Par ailleurs, nous avons choisi dans ce travail de présenter les échantillons en monadique pur dans un souci de praticité et de façon à raccourcir la durée de la séance. Afin de valider l'intérêt de cette méthode, il serait maintenant intéressant de réaliser le même travail avec des recettes plus différentes sensoriellement en présentant les produits en monadique séquentiel ce qui nous permettrait de nous affranchir d'un effet *sujet* et de peut-être obtenir des résultats plus significatifs. Nous pourrions ainsi vérifier la capacité de la méthode à révéler de fortes interactions *produit* \times *situation* particulièrement riches et peu prises en compte aujourd'hui dans les études consommateurs.

C – LES APPROCHES ORIENTEES « IMPLICATION »

Dans cette deuxième partie, nous avons souhaité nous attaquer au problème de la faible implication des participants dans leurs réponses lors d'un test en salle. En effet, nous pensons que lors du test à domicile, les sujets s'engagent à consommer les produits pendant une certaine période et ont donc le temps de se faire une opinion sur le produit en étant impliqué. Bien évidemment, les méthodologies présentées précédemment basées sur la prise en compte des conditions de consommation des produits lors des tests en salle améliorent sûrement l'implication du sujet dans la dégustation des produits. Cependant, nous souhaitons également nous intéresser à l'amélioration de l'implication du sujet dans sa réponse directement par le biais de la tâche expérimentale.

Nous avons choisi d'explorer deux types de mesures indirectes pour révéler les performances organoleptiques des produits sur la base de réponses a priori plus impliquées :

- le degré d'authenticité perçue d'un produit,
- la propension à échanger un produit contre un autre.

Afin d'obtenir un panorama complet de l'ensemble des propositions méthodologiques sur les mêmes produits, nous avons mené ces deux investigations méthodologiques avec les deux recettes de biscuits feuilletés au fromage précédemment comparées et pour lesquelles le test en salle traditionnel n'a pas permis d'obtenir les mêmes résultats que le test à domicile, à savoir une meilleure performance de la recette A. Nous espérons donc, par l'intermédiaire des deux protocoles envisagés, arriver à mettre en évidence la supériorité de la recette A sur la recette B.

Ce chapitre est scindé en deux parties qui abordent séparément les deux approches méthodologiques testées pour augmenter l'implication du participant dans la tâche expérimentale. Chacune des deux parties présente le principe général de l'approche méthodologique ainsi que l'application du protocole, sous forme d'article, pour la comparaison de l'appréciation des deux recettes de biscuits feuilletés au fromage.

I - LE TEST D'AUTHENTICITE

I.1 - PRINCIPE DE LA METHODE D'AUTHENTICITE

L'amélioration de l'implication du participant dans la réponse recueillie est loin d'être aisée car elle ne doit en aucun cas se faire au dépend du maintien de l'état affectif du participant. En effet, d'après plusieurs auteurs (Kihlstrom, 1987 ; Shiv et Fedorikhin, 1999 ; Zajonc, 1980, 1984) nos décisions rapides (ce qui est le cas du choix d'aliments) ne sont pas toujours prises avec une intervention consciente. Par ailleurs, les travaux de Dickson et Sawyer (1990) ayant examiné la façon dont les consommateurs prennent réellement des décisions dans différents points de vente suggèrent que les décisions d'achat sont plus souvent guidées par les sentiments que par la pensée. En ce qui concerne le comportement alimentaire, l'analyse des raisons de préférence est un acte que nous réalisons finalement peu au quotidien et les préférences n'ont, en aucun cas, besoin de processus cognitif conscient pour se former de manière naturelle. Une attitude analytique basée sur les avantages et inconvénients de différentes alternatives pourrait même aboutir à des choix qui ne seraient finalement pas optimaux c'est-à-dire satisfaisants (Wilson et Dunn, 1986 ; Wilson et Schooler, 1991). Compte tenu de cet aspect, augmenter l'implication du participant dans la tâche ne doit en aucun cas l'amener à être plus analytique envers les produits qu'il déguste puisque ce n'est pas quelque chose qu'il fait de manière aussi forte et consciente dans la vie de tous les jours.

Ce compromis entre un bon niveau d'implication des sujets tout en permettant une analyse affective des produits est ce que propose la méthode du test d'authenticité développée par Mojet et Köster (1986). L'idée est de récupérer une information hédonique par l'intermédiaire d'une question indirecte basée sur l'identification d'un produit comme étant authentique ou faux. La question sur l'authenticité du produit véhicule pour le sujet une notion de performance dans la tâche et est donc, à elle seule, plus engageante que la simple question d'appréciation. Par ailleurs, cette tâche d'identification pouvant favoriser la mise en place de processus cognitifs, il convient de minimiser l'attitude analytique créée chez l'individu par l'environnement formel du test en le projetant dans une attitude émotionnelle à l'égard des produits. Les réactions émotionnelles et impliquées sont donc motivées par une étape d'amorçage négatif affectif pendant laquelle on explique au participant l'enjeu du test pour l'industriel qui est de modifier la fabrication du produit authentique dans le but d'augmenter le profit de l'entreprise au dépend du consommateur. Le comportement éthique d'une entreprise a effectivement été démontré comme étant particulièrement impliquant dans

le comportement d'achat des consommateurs (Creyer et Ross, 1997). Ainsi, ce sentiment « d'arnaque » rendrait les sujets particulièrement critiques ce qui libèrerait leurs sentiments immédiats qui ne seraient alors plus sous contrôle rationnel et analytique. Par ailleurs, l'amorçage d'un état émotionnel négatif chez des participants les pousserait à être plus performants lors d'une tâche de détection (Chang et Wilson, 2004). Après cette première étape d'amorçage émotionnel, le protocole consiste à présenter au participant, plusieurs répétitions des différents produits et à lui demander s'il pense que l'échantillon dégusté correspond à la recette authentique du produit ou non.

I.2 - APPLICATION DE LA METHODE POUR COMPARER LES PERFORMANCES ORGANOLEPTIQUES DE DEUX BISCUITS FEUILLETES AU FROMAGE (ARTICLE 3)

Bien que la méthode du test d'authenticité ait été initialement développée pour évaluer des préférences (Mojet et Köster, 1986), son utilisation a été rapportée plutôt pour améliorer la discrimination de produits proches sensoriellement (Frandsen *et al.*, 2003, 2007). Par ailleurs, ces études mettent généralement en jeu des catégories de produits consommés régulièrement comme du lait ou de la bière pour lesquels la notion d'authenticité a un réel sens. Enfin, les sujets interviewés sont généralement des consommateurs très réguliers des produits testés et ont donc des attentes claires en ce qui concerne le vrai goût du produit grâce à la mémorisation implicite de leur référence interne.

Compte tenu du cadre matériel dans lequel nous souhaitons utiliser cette méthode (c'est-à-dire avec des produits industriels et des sujets non-consommateurs d'une marque bien spécifique), nous avons réalisé une phase de pré-test avec 30 sujets qui nous a permis de vérifier que cette méthode du test d'authenticité pouvait être mise en place :

- avec une catégorie de produits comme des crackers salés pour laquelle le lien avec la notion d'authenticité n'est pas évident (production industrielle),
- avec une présentation répétée des échantillons malgré deux recettes légèrement différentes visuellement et donc facilement reconnaissables,
- avec des sujets consommateurs de la catégorie de produits (biscuits feuilletés salés) mais pas forcément de la recette testée qui ne possèdent donc pas forcément une référence interne exacte.

Cette phase de pré-test a également été l'occasion de tester plusieurs scénarios d'amorçage et d'en choisir un qui soit :

- crédible sur des produits de fabrication industrielle,
- efficace pour amorcer un état émotionnel négatif chez le participant.

Enfin, nous avons pu vérifier la possibilité d'appliquer cette méthode lors d'un test en salle avec un recrutement ponctuel en rue des sujets. Le protocole de test auquel nous avons abouti est donc :

- applicable lors d'une séance de dégustation de vingt minutes maximum,
- facile à mettre en œuvre par des enquêteurs non expérimentés,
- facile à comprendre et à réaliser par les participants.

L'article qui suit présente la mise en application du protocole de test d'authenticité développé pour comparer les performances organoleptiques des deux recettes de crackers salés au fromage.

ARTICLE 3 : Boutrolle I., Delarue J., Köster E.P., Arranz D., Danzart, M.

How to improve respondents' commitment when evaluating food acceptance: use of the authenticity test. *Food Quality and Preference*. Soumis en octobre 2006.

RESUME

Les développeurs de produits alimentaires vérifient fréquemment l'appréciation de leurs recettes en demandant aux consommateurs de déclarer leurs préférences entre plusieurs échantillons. Cette approche est souvent critiquée pour le manque d'engagement des participants dans la réponse hédonique. Cette étude a pour objectif de tester la possibilité d'utiliser la méthode d'authenticité comme alternative à un test hédonique traditionnel pour l'évaluation de l'appréciation de deux recettes de crackers salés. Cette méthode présente l'avantage de fournir des réponses affectives et non analytiques par le biais d'une augmentation de l'implication des participants dans le test. Un groupe de 244 femmes a été interviewé selon le design d'un test d'authenticité. Une attitude affective et négative envers les échantillons à évaluer est induite chez les participantes en leur racontant l'histoire suivante : « *Nous souhaitons tester la possibilité de remplacer la version originale de notre marque de biscuit feuilleté par une copie fabriquée avec des ingrédients moins coûteux. Bien entendu, le prix de vente restera identique même si les ingrédients sont moins chers* ». Les participantes doivent ensuite déguster plusieurs échantillons (6 répétitions de chaque produit) et les identifier comme étant une version originale ou une copie du biscuit. A la fin de la séance, les participantes reçoivent une nouvelle paire des deux produits et doivent alors identifier l'échantillon qu'elles préfèrent. Les résultats du test d'authenticité montrent que globalement les réponses d'authenticité sont en adéquation avec les préférences. Cependant, la double analyse authenticité/préférence a permis d'aboutir à des informations particulièrement riches sur la diversité de la perception des consommatrices. Par ailleurs, les réponses à la question de préférence posée lors du test d'authenticité sont comparées à celles obtenues dans deux tests hédonique traditionnels mis en place dans les conditions d'un test en salle et d'un test à domicile.

Mots clés : *Test d'Authenticité, Test de Préférence, Implication, Amorçage, Théorie de la Détection du Signal, Exposition répétée.*

ABSTRACT

Food developers frequently check the liking for their recipes by asking consumers to state their preferences. This approach is often criticised for the lack of commitment of the participants in the hedonic response. This study tested whether an authenticity test could be used as an alternative to a traditional hedonic test when evaluating the liking for two brands of salted crackers. Such a method has the advantage that responses are affective and not analytical and that the participants' involvement is expected to be enhanced. A group of 244 women were interviewed according to an authenticity test design. A negative affective attitude toward the samples to be evaluated was induced by telling the participants the following upsetting story: "*We want to investigate the possibility of replacing the genuine version of our salted crackers brand by a copy manufactured with low-cost ingredients. Of course, the retail price will be the same, even though the ingredients are cheaper*". The participants were then asked to say whether each presented sample (6 repetitions of each product) was a genuine one or a copy. At the end of the session, subjects received a new pair of samples and were asked to say which one they preferred. Responses to this preference question were compared to those obtained with traditional hedonic tests run in Central Location Test and Home Use Test conditions. Results of the authenticity method show that overall authenticity responses are closely linked to preferences. Besides, these paired authenticity/preference data provide insightful information on the diversity of consumers' perception.

Keywords: *Authenticity test, Preference test, Involvement, Priming, Signal Detection Theory, Repeated exposure*

HOW TO IMPROVE RESPONDENTS' COMMITMENT WHEN EVALUATING FOOD ACCEPTANCE: USE OF THE AUTHENTICITY TEST.

Isabelle Boutrolle^{a,b}, Julien Delarue^b, Ep Köster^c, Delphine Arranz^a, Marc Danzart^b

^a *Sensovaleur-Danone Research, RD 128, 91767 Palaiseau cedex, France*

^b *Laboratoire de Perception Sensorielle et Sensométrie, UMR SCALE (ENSIA-CNAM-INRA), 1 avenue des Olympiades, 91744 Massy Cedex, France.*

^c *Wageningen University and Research, Center for Innovative Consumer Studies, Bornsesteeg 59, 6708PD Wageningen, The Netherlands.*

1. INTRODUCTION

Food developers and sensory scientists must frequently check the performance of their recipes. In such a case, quantitative market research companies usually propose the implementation of hedonic blind tests. Among the existing methodologies, the Central Location Test (CLT) design remains the most popular method and represents around 70% of all hedonic blind tests implemented in France. The CLT consists of an evaluation of the products by a large consumer panel after a brief exposure under standardised conditions of consumption. The Home Use Test (HUT), another well-known technique, is also sometimes implemented to assess the liking for food products. It consists of a repeated use of the product under common daily use circumstances, usually for a period of several days. Unfortunately, the HUT, which might be assumed to better predict the actual consumption behaviour, is seldom used due to logistic (i.e. short shelf life, difficulties to package blind) and budget constraints. The main criticism raised against CLT is that the tasting conditions are very different from natural eating situations and do not allow the respondents to be emotionally involved like in actual food consumption. Hence, there is every chance that these artificial tasting conditions will bring about casual responses and will induce an analytical attitude in the subjects.

The aim of the present paper was to develop a CLT protocol combining the ease of use of the CLT with an affective commitment of the subjects as in real life. We tested whether an authenticity test could be used as an alternative to a traditional hedonic test. The principle of such a test, which was initially developed for difference testing by Mojet and Köster (Köster, 1990; Mojet & Köster, 1986), is to induce a critical affectively negative attitude in the

participants. This is achieved by telling them prior to the evaluation that the firm can raise its profit by making low-cost versions of their favourite product while selling it at the same price. Participants are then presented with a series of samples and are asked to say for each of them whether it is a genuine one or a copy. The advantage of such a measurement is that the consumers are oriented toward an emotional, non-analytical attitude, which hopefully will lead to judgements based on affective rather than analytical processes. This is in accordance with Zajonc's theory (1980) who claimed that affective reactions are immediate and occur without cognitive processing. In a previous experiment conducted with two brands of salted crackers, we observed that the CLT and the HUT led to different results. The HUT appeared to be more discriminative, leading to a significant difference in liking between the two types of tested crackers, whereas the CLT did not. Hence, it seems interesting to check with these two products if the emotional mindset that we induced in the participants using the authenticity method will lead to results that are closer to the HUT results than those of the CLT are. Furthermore, we investigate the relevance of using the degree of authenticity as an indirect measure of liking by assuming that the participants answer the authenticity question according to their hedonic sensory expectations. Eventually, we discuss the interest of using authenticity data together with traditional hedonic data.

2. MATERIALS AND METHODS

2.1 SAMPLES

Two brands of salted cheese crackers, a national brand and a private label were tasted and evaluated according to the authenticity test design. The sensory profiles carried out on the two types of cheese crackers revealed many differences between these two products in terms of visual, texture and taste attributes. No information was given to the subjects regarding the products (brand, price...). Samples were presented in aluminum cups that each contained three units of crackers, allowing the participants to account for the small within-batch variability.

2.2. SUBJECTS

Two hundred and forty four females between 18 and 65 year old participated in the experiment (age 41.8 ± 12.8). They were recruited in three different French cities. All subjects usually consumed salted crackers at least several times a month and had no aversion to cheese flavour. Their preferred brand of cracker was asked, but was not used as a selection criterion.

2.3. PROCEDURE

2.3.1 Priming sequence

Subjects were recruited in the street throughout the day and were invited to a testing room to participate in a 15-min face-to-face interview. First, in order to induce a negative attitude, the interviewer told the subject the following upsetting story: “We want to investigate the possibility of replacing the original version of our salted crackers brand by a copy manufactured with low-cost ingredients. This will be only possible if the consumers accept the low-cost product and if it tastes like the original version. Of course, the retail price will be the same, even though the ingredients are cheaper”. The story was read aloud to the participants who also had a written version of the text in front of them during the task. By referring to low-cost ingredients, we expected that, rather than focusing on a specific sensory alteration, the subjects would evaluate the samples as if they had undergone an overall modification.

2.3.2. Authenticity task

After they had been told the story, the consumers had to identify each item of a series of crackers samples as either ‘original’ or ‘copy’. First, a warming up sample, either the national brand or the private label brand, was presented. Then, each subject had to sequentially taste and identify 12 samples. These 12 samples consisted of 6 national brand samples and 6 private label samples presented in a specific order. In order to account for possible carry-over effects, the presentation order of the samples was designed to balance the following sequences over the individual successive sample presentations: AA, BB, AB and BA. In addition, for each evaluated sample the subjects had to rate their degree of confidence in their responses on a five point scale ranging from 1 (“I am not sure at all about my response”) to 5 (“I am sure about my response”).

Finally, at the end of the session, subjects were sequentially presented with a new pair of samples while specifying to them that one of these two samples was the original version and the other was a copy. Here, the subjects were asked to say which was their preferred sample. A “no preference” response was admitted.

2.4 DATA ANALYSIS

2.4.1. Authenticity task

Responses to the warm-up sample were excluded from the analysis. Individual responses were structured according to the theory of signal detection (SDT) as illustrated in Table 1, defining hits (HIT) as the ‘original’ answers when the national brand was presented and false alarms (FA) as the ‘original’ answers when the private label product was presented. Consequently, the misses (MISS) were defined as ‘copy’ answers to the presentation of national brand and the correct rejections (CR) as ‘copy’ answers’ to the private label presentation.

Table 1. Structure of the individual data according to the SDT.

		Presented sample	
		National Brand	Private label
Response type	‘Original’	HITS	FALSE ALARMS
	‘Copy’	MISSES	CORRECT REJECTIONS

An overall table of “subject × SDT” responses was obtained and individual d' values were calculated (Frandsen, Dijksterhuis, Brockhoff, Nielsen & Martens, 2007). Given the fact that the definition of HIT and FA is arbitrary, the d' calculation can lead to negative or positive values. Thus, a positive d' value indicates that the subject considered the national brand as the ‘original’. In contrast, a negative d' value indicates that the subject considered the private label to be the ‘original’. It is commonly admitted in sensory difference testing that a d' value above 0.5 indicates that the samples are successfully discriminated.

Overall data

To conclude whether the panel discriminates the two samples or not, the mean of the absolute individual d' values was calculated and the rejection of H_0 : mean $|d'| = 0$ was tested using a t-test ($\alpha=5\%$). Besides, the mean of the individual d' values was also calculated and tested against zero using a t-test to evaluate the overall direction of the authenticity difference between the two products.

In addition, the degree of authenticity of each brand was determined by calculating the proportion of ‘original’ responses to each of the two products on all the evaluations.

Confidence ratings were also analysed to get some information on the way the participants perceived the task. The influence of the tasted product and of the type of response

(i.e. original or copy) on participants' confidence was tested by ANOVA using the following model: $confidence\ score = product + response\ type + product \times response\ type$.

Participant typology

In order to better take into account the diversity of individual responses, a cluster analysis was performed on the "subject \times SDT" table. The calculation was based on Euclidean distance and Ward's criterion. This analysis allows to identify groups of consumers who perceived the genuineness of the products in similar ways.

Dynamic analysis

A χ^2 test ($\alpha=10\%$) was carried out for each successive sample presentation in order to observe the evolution of the sample discrimination over exposures.

2.4.2. Preference question

The percentages of subjects who preferred the national brand and those who preferred the private label were compared using a Binomial test ($\alpha = 5\%$). Authenticity data were then re-analysed in dependence of the subjects' stated preference at the end of the session to explore the link between authenticity and preference data.

Overall consistency

First, to test the assumption that the preferred product is the product which is identified as 'original', a modified "subject \times SDT" table was constructed, defining Hits as the 'original' answers when the preferred product was presented and FA as 'original' answers when the less liked product was presented. Thus, new individual d' values were obtained. Subjects who reported having "no preference" were excluded from the data set. The link between the degree of authenticity and preference can also be investigated by comparing the proportions of 'original' responses for each product within two groups: subjects preferring the national brand and subjects preferring the private label.

Individual consistency

Finally the subjects' "consistency" was checked by comparing for each participant which product was preferred and which was considered to be the 'original' as revealed by the analysis of the sequence of individual responses during the authenticity task.

3. RESULTS AND DISCUSSION

3.1. METHODOLOGICAL CONSIDERATIONS

The duration of the session was less than 20 minutes, which is quite good since it is similar to the duration of a traditional CLT session. However, because of the individual sample presentation order, the preparation of the session was a little bit more complicated to manage. Furthermore, it seemed that the negative priming worked well for the majority of participants with some very negative reactions (one woman refused to participate in the study once she heard the upsetting story). However, participants were sometimes happy that, for once, we “told them the truth”. During the authenticity task, the repeated tasting was not really a problem, even if we observed of course some lassitude at the end of the sequence. Nevertheless, subjects tasted at least one cracker unit for each presented sample.

3.2. AUTHENTICITY TASK RESULTS

Authenticity data were gathered and analysed as it is usually done for common hedonic data, i.e. first for the whole panel and then on an individual basis.

3.2.1 Overall data

Degrees of authenticity results

When taking all samples and all subjects into consideration, the test leads to 40% of ‘original’ responses. The participants indeed answered ‘copy’ more often than ‘original’. This phenomenon could be an illustration of the uncertainty of the subjects in answering such a question or might reflect a negative feeling induced by the upsetting story placing the subjects in a suspicion mindset.

The analysis of the overall data set according to SDT shows that the mean of the absolute individual d' values is significantly different from zero (mean $|d'| = 0.94$; $t = 19.40$; $p < 0.001$). This indicates that the consumers were able to perceive an authenticity difference between the two crackers, irrespective of the direction of the difference was.

However, it is more informative to compare the overall degree of authenticity of the two separate brands. As expected on the basis of the results of the previously executed HUT, the test invoked a higher percentage of ‘original’ responses to the national brand (42%) than to the private label (36%). The overall mean of the individual d' values (mean $d' = 0.17$; $t =$

2.20; $p = 0.028$) which is positive and significantly different from zero shows that the national brand was seen as more authentic than the private label. However, this d' value is far from the threshold of 0.5 indicating a slight overall difference of genuineness between the two crackers. The fact that there is a large difference between the mean of the absolute values of the individual d 's and the mean of the positive and negative individual d 's indicates that there is probably a bimodal distribution in the data. Segmentation of the subjects on the basis of this will be discussed below.

Confidence ratings results

Confidence ratings provide some insight in the way the consumers express themselves. Figure 1 shows the average confidence ratings for the four combinations of the response types (original or copy) and the stimuli (national or private brand).

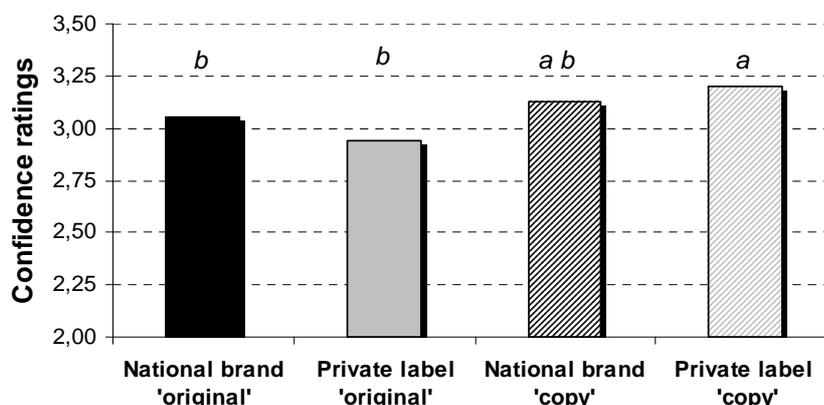


Fig. 1. Mean confidence ratings to the response types upon tasting each of the two products. (similar letters indicate that the means are not significantly different as tested by ANOVA, $\alpha = 0.05$).

As can be seen, the subjects were more confident about their 'copy' responses than about their 'original' responses, irrespective of the tasted sample. ANOVA carried out on the confidence ratings shows that the *response* effect is clearly significant ($F_{1,3105} = 15.63$; $p < 0.001$). At the same time, the *stimuli* \times *response* interaction is significant ($F_{1,3105} = 6.50$; $p = 0.011$) with a higher confidence for 'copy' responses to the private label than to the national brand and a higher confidence for 'original' responses to the national brand than to the private label stimuli. This is consistent with the trend to a better identification of the national brand as being the "original".

3.2.2. Subgroup analysis

Cluster analysis results: degrees of authenticity according to the response pattern

The difference in degree of authenticity between the national brand and the private label is not large enough to conclude that the national brand is seen as more authentic. This is due to a clear segmentation of the respondents. Some participants identified the national brand as the “original” whereas others identified the private label as the “original”. Figure 2 shows the result of the cluster analysis applied to the data of the SDT table. This classification leads to four different groups of consumers. Subsequent cluster analyses carried out on the response patterns of each of these groups reveal how these consumer groups differ.

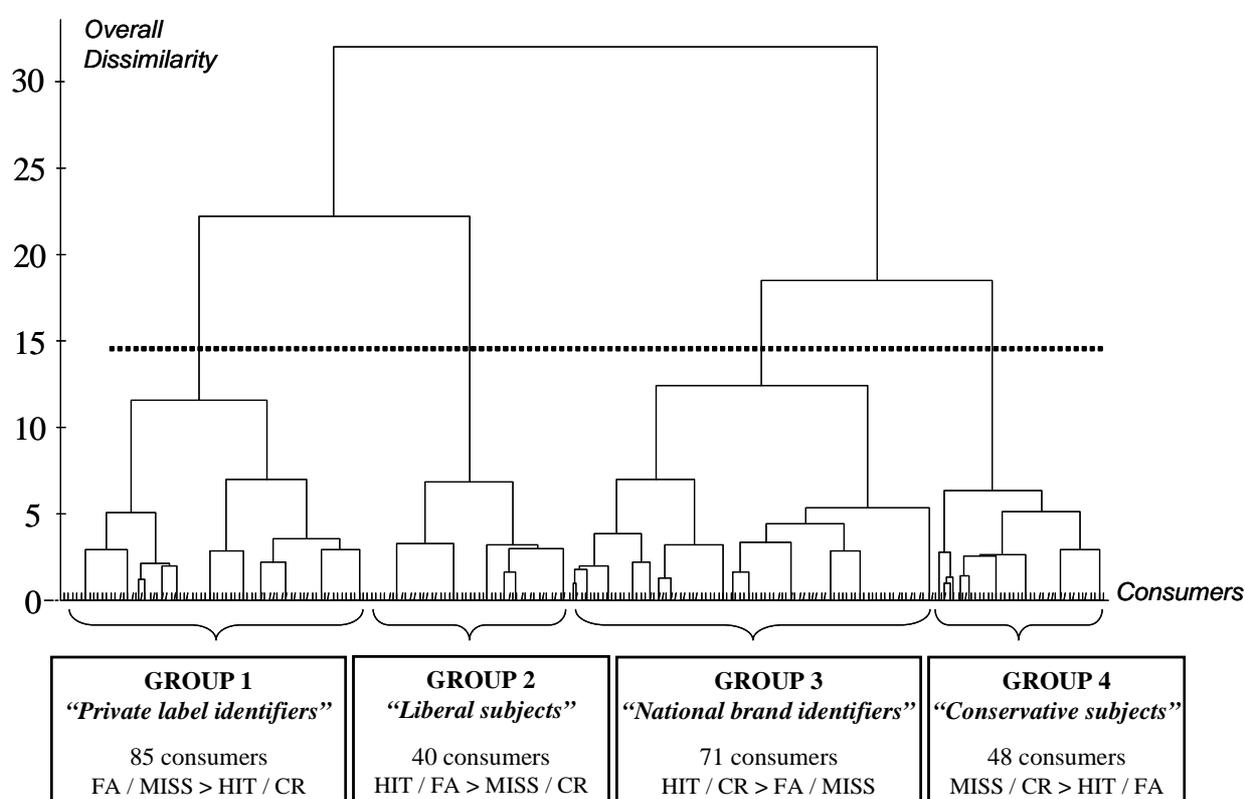


Fig. 2. Cluster analysis results: consumer groups and associated response patterns.

Group 1 (85 subjects) consists of respondents who reported many more FA and MISS responses, indicating that they have almost certainly used the private label as their reference for the guessed original product. In contrast, the response pattern of the consumers in group 3 (71 subjects) shows more HITs and CRs, indicating that these subjects have identified the national brand to be the original product. The two other groups encompass subjects who tended to give either more ‘original’ responses (group 2: 40 subjects) or more ‘copy’

responses (group 4: 48 subjects) regardless of the sample. These two groups differ in the decision criterion (τ) they used, group 4 (mean $\tau = 0.95$, SD = 0.19) using a stricter one than group 2 (mean $\tau = -0.32$, SD = 0.28). According to the terminology of the signal detection theory, participants who tend to give more 'original' responses are defined as "liberal subjects" and those who tend to give more 'copy' responses are defined as "conservative subjects".

Figure 3 presents the proportions of 'original' and 'copy' responses according to the presented sample for each group. We then calculated intra-group mean d' in order to evaluate the degree of differentiation between the two products in terms of authenticity.

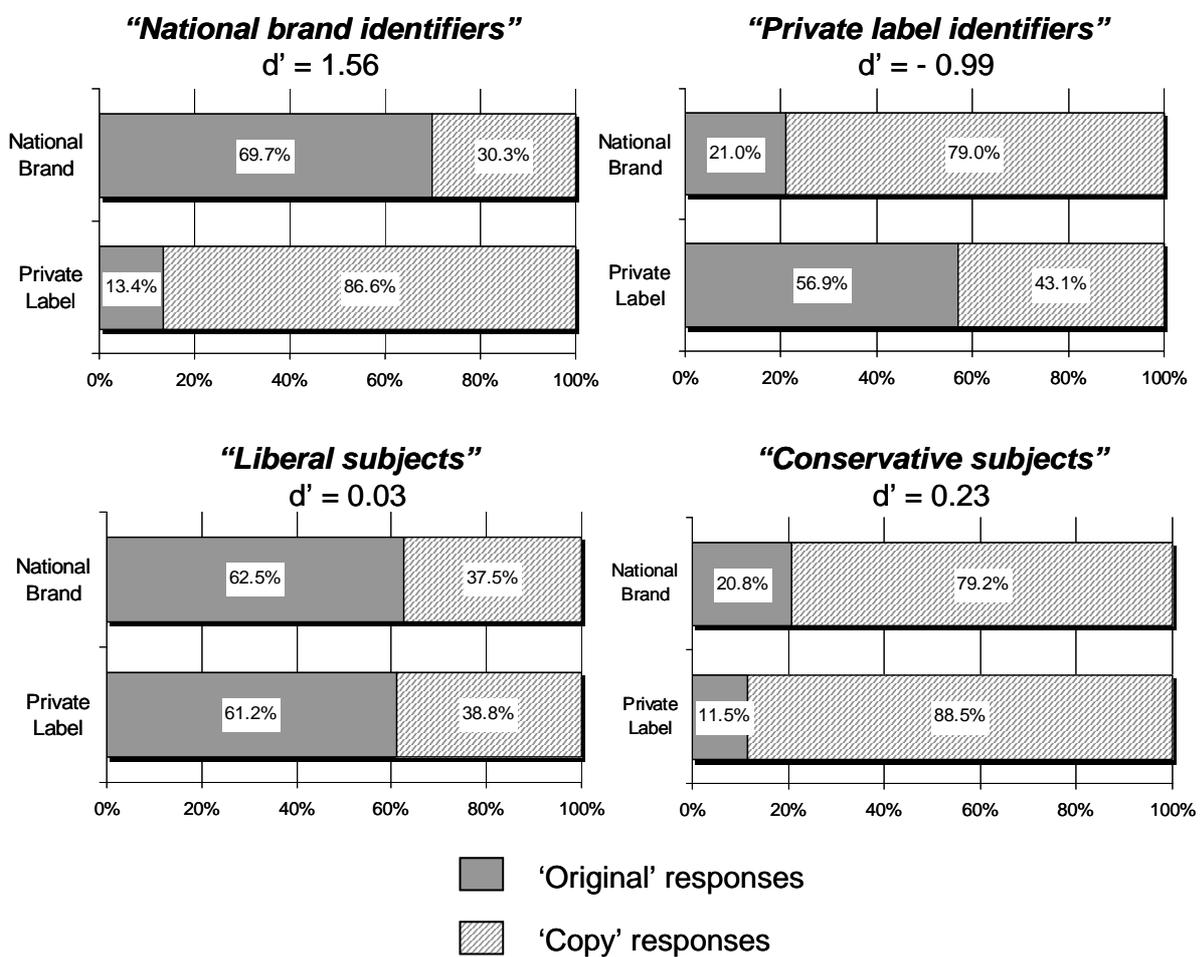


Fig. 3. Percentages of 'original' and 'copy' responses according to the tasted sample for each consumer group.

As can be seen in figure 3, the mean d' of the "private label identifiers" (group 1) is negative (mean $d' = -0.99$; $t = -13.85$; $p < 0.001$), whereas the mean d' calculated for the "national brand identifiers" (group 3) leads to a positive value (mean $d' = 1.56$; $t = 18.30$; $p < 0.001$). The mean $|d'|$ of the "national brand identifiers" is significantly higher than the one

of the “private label identifiers” ($t = 5.062$; $p < 0.001$). This indicates that the subjects who identified the national brand as the authentic sample are more discriminating than the subjects who identified the private label as the authentic one. Thus, the national brand is perceived as more authentic for the “national brand identifiers” than the private label for the “private label identifiers”. In other words, even if more participants used the private label as the authentic sample than the national brand, the genuineness of the national brand could be seen as more stable than the genuineness of the private label.

As far as the two other groups are concerned, the low mean d 's confirm that these consumers perceived less difference of authenticity between the two products (“liberal subjects”: mean $d' = 0.03$; $t = 0.359$; $p = 0.722$; “conservative subjects”: mean $d' = 0.29$; $t = 4.38$; $p < 0.001$). It is worth noting that one third of the overall interviewed subjects did not report a difference of authenticity between the two products. However this does not necessarily mean that they do not perceive a difference between the products. Indeed, this might also reflect a certain lack of confidence of the subjects while performing the task, resulting either in casual or in over-suspicious response behaviours.

It is also very informative to explore the performance of the national brand users with regard to the identification of their favourite product. We expected that the subjects who had extensively experienced the national brand product before would have performed better in the identification of the national brand as the authentic product. In fact, the 58 subjects who spontaneously reported that their usual brand of salted crackers is the national brand did not easily identify their favourite brand. Although the users seem to be a little bit more discriminating than the non-users, this difference is not significant. Actually, users and non-users present the same overall response patterns. However, even if the recognition performance of the national brand users is rather poor, they are significantly more confident in their responses than the non-users ($t = 7.137$; $p < 0.001$).

Confidence scores results for each consumer group

Figure 4 shows the confidence scores for each group of consumers. The confidence pattern of the “national brand identifiers” illustrates an expected behaviour with the highest confidence found for “Correct Rejections” (the private label identified as ‘copy’). This is in line with previous studies where the authors (Köster, Prescott & Köster; 2004; Møller, Wulff & Köster, 2004; Mojet & Köster, 2005) observed that, even if consumers have difficulties detecting what they have already experienced, they perfectly detect what they had never experienced. In the same way, the “private label identifiers” who used the private label as

their reference, present the same confidence pattern with the highest confidence for the national brand identified as 'copy'.

Interestingly, we can see that "liberal subjects" are overall more confident in their responses although they did not discriminate the samples. This probably indicates that they did not care about the difference of authenticity even if they perceived it. Perhaps the emotional priming sequence did not work with these levity participants who thus reported casual responses. Conversely, confidence scores of the "conservative subjects" reveal that these participants reported their numerous 'copy' responses with high confidence regardless of the stimuli. Accordingly, the low confidence scores for the few 'original' answers reflect their conservative and suspicious behaviour. The negative affective priming sequence may have excessively worked for these participants with the result that they did not want to risk identifying a false sample as an original one and were thus reluctant to report that a sample is an original version. In other words, their decision criterion was very strict. They thus reported many 'copy' responses leading to non-discriminating results. Nevertheless, we think that these participants were certainly more involved than the "liberal subjects" and tried hard to perform the task.

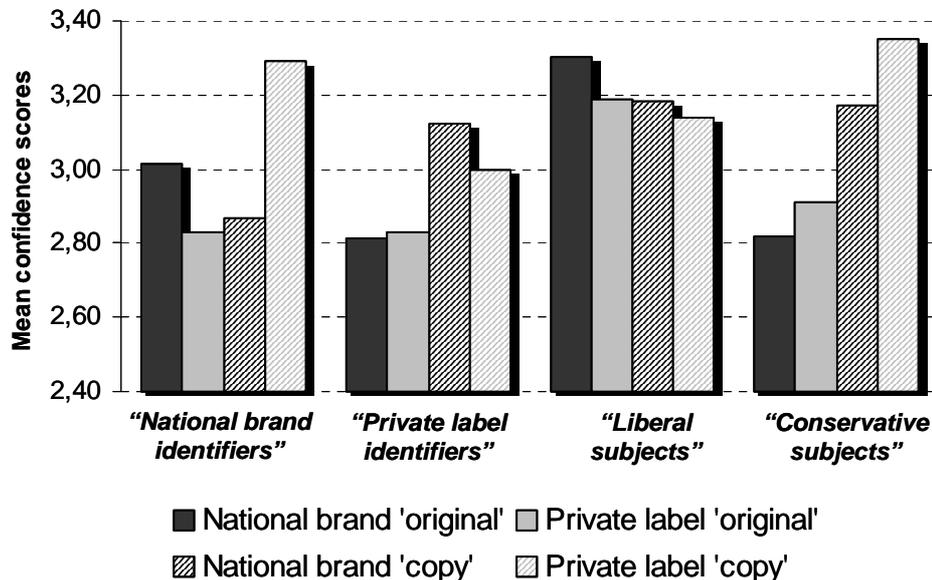


Fig. 4. Mean confidence ratings according to the tasted product (national or private brand) and to the response type ('original' or 'copy') for each consumer group.

3.2.3 Dynamic analysis

Overall data

Figure 5 represents the evolution of the percentage of 'original' responses over the sample and the presentation sequence. This figure shows that the subjects more frequently identified the warm-up sample as 'original' (above 50% of the responses) than the following samples (around 40% of the responses). Furthermore, χ^2 analyses carried out for each successive sample presentation show that the difference between the degrees of authenticity (i.e. percentage of 'original' responses) of the two products is hardly ever significant. Nevertheless, it is clear that the percentages of 'original' responses on all occasions are higher for the national brand.

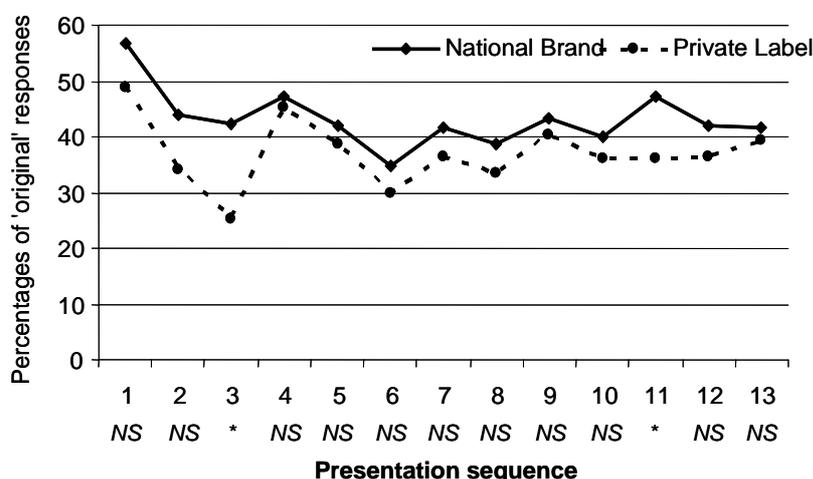


Fig. 5. Percentages of 'original' responses according to the product and to the presentation sequence. χ^2 results are indicated for each position (*: significant difference ($p < 0.05$); NS: non significant difference).

Subgroup data

Figure 6 shows the development of the percentages of 'original' responses according to the sample and to the presentation sequence for each group of consumers.

The comparison of the graphs illustrating the responses of the two groups of discriminating subjects (national brand and private label identifiers) confirms the previously cited results. Indeed, it can be clearly seen that those who identified the national brand as the authentic product are more discriminative than participants who used the private label as their reference, including for the first sample. We can also observe a clear impact of the sample presentation order on the percentages of 'original' responses. For the national brand tasted by the "private label identifiers", we observe that the percentage of 'original' responses is the highest when the national brand sample is presented as the first one in the series. In a similar

way, the results of the “national brand identifiers” show that the percentage of ‘original’ responses for the private label samples is the highest when it is presented for the first time. Thus, the genuineness of the two samples is over-evaluated when they are tasted for the first time. Interestingly, this phenomenon looks like the well-known “over-scoring” for the first tasted product in hedonic evaluations (Stone & Sidel, 1993). Therefore, this observation highlights a similarity between the respondents’ behaviour when answering a hedonic question and when answering an authenticity question. The most striking illustration of the presentation order effect concerns the “conservative” subjects who were initially less conservative with the first samples. Interestingly, even if these participants did not perceive a significant difference of authenticity between the two products, the percentages of ‘original’ responses are higher for the national brand on almost all occasions. On the contrary, the response pattern of the “liberal subjects” reveals their uncertainty concerning the identification of the authentic product over the entire presentation sequence.

Overall, we can see that for all groups, the main change takes place over the first two or three samples. One might consider whether in future studies, it might be possible to complete the task with fewer samples. On the other hand, this might restrict the use of the individual data as a source of information.

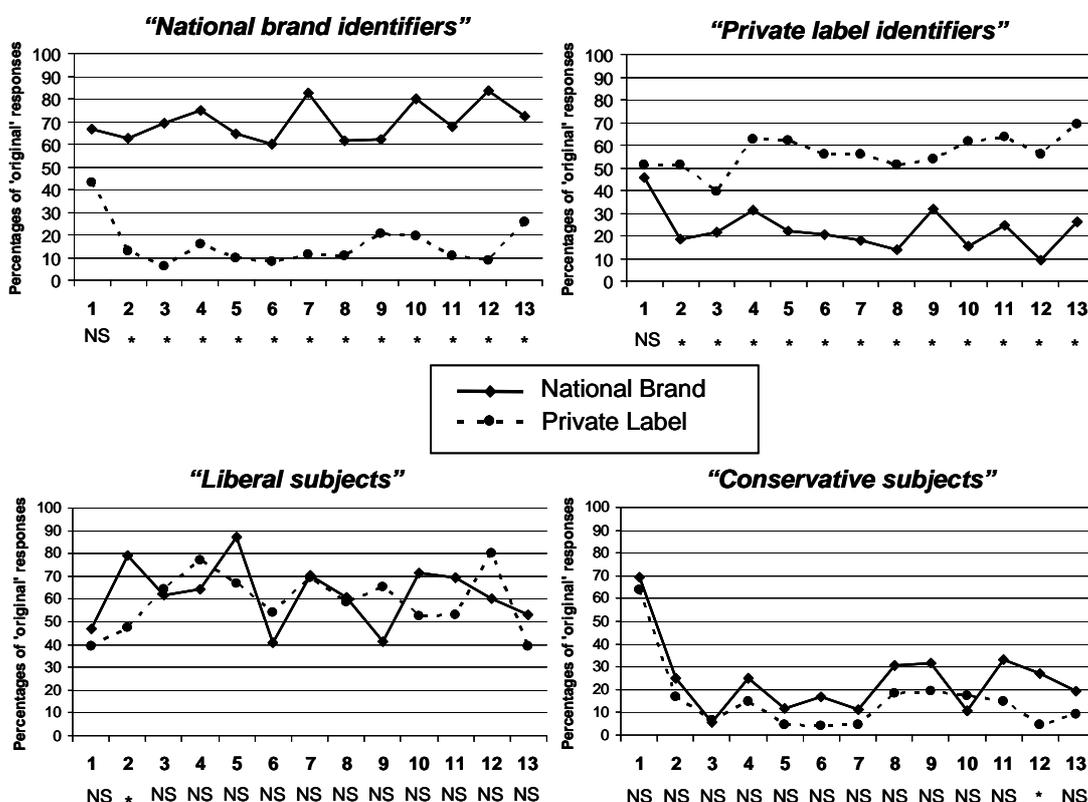


Fig. 6. Percentages of ‘original’ responses according to the product and to the presentation sequence for each group of consumers. χ^2 results are indicated for each position (*: significant difference ($p < 0.05$); NS: non significant difference).

3.3. Reported preferences

3.3.1 Preference results

The preference question asked at the end of the session showed that, among the subjects who reported a preference toward one brand, significantly more subjects ($p = 0.015$) stated a preference for the national brand (57.6%) than for the private label (42.4%). Besides, 27 subjects reported having “no preference”. The preference question of the previously conducted traditional CLT did not reveal such a high preference for one brand. This higher discrimination observed in the present experiment is promising, should one wish to reveal slight preferences. Frandsen, Dijksterhuis, Brockhoff, Nielsen, and Martens (2003) already found that the authenticity test could be more discriminating than descriptive analysis.

Interestingly enough, preference results are not linked to the user status. Actually, 58.8% of the national brand users stated they preferred the national brand. This proportion was approximately the same for the non-users (57.2% of national brand preferrers). Did some users recognize their usual product but finally preferred the competitor product? As previously mentioned, the authenticity results according to the user status did not reveal such behaviour. A more likely and simpler explanation is that numerous users did not easily identify their usual product revealing that the sensory identity of the national brand product would not be so well memorized by these supposedly loyal consumers.

The hedonic information of the combined authenticity-preference test is similar to HUT hedonic conclusion where 65.5% of the subjects stated that they preferred the national brand. The authenticity-preference test output is thus more similar to HUT output than the traditional CLT output. Several explanations could account for the better performance of the authenticity protocol to obtain the HUT conclusion. First, the specific repeated exposure design gives access to judgements (either authenticity or preference) that are based on several consumptions of the tested products, similarly as in HUT where the subjects ate at least 1 pack of crackers (100g). On the contrary, during the traditional CLT, the subjects ate only few units of crackers (and rarely more than 1 unit) before stating their preference. Numerous authors have shown the influence of repeated exposure on hedonic measures (Lévy & Köster 1999; Lévy, MacRae, & Köster 2006; Meiselman, De Graaf, & Leshner, 2000). It is thus likely that such a brief exposure in CLT creates a false impression leading to a judgement that differs from the one obtained after repeated exposures in the HUT and the authenticity-preference test. Besides, it is very likely that the protocol design affects the subjects' internal state and thus their involvement in the evaluation task. We assume that the more natural and

time-consuming conditions (1 week) of the HUT protocol enhance participants' involvement. In contrast, CLT implementation may tend to induce a lack of involvement of the participants in the test leading to casual responses. In the present experiment, we observed that the subjects were very concentrated on the task during the authenticity-preference test. It appears that the authenticity task tends to increase the subjects' involvement during the whole task until the end of the test including the preference question.

3.3.2 Relationship between authenticity and preference: overall data

In order to investigate the link between the authenticity and the preference, the authenticity data from the initial sequence were analysed in the light of the subjects' stated preference at the end of the session. We calculated new individual d' values on the basis of the modified "subject \times SDT" table now defining the HIT as the 'original' answer when the preferred product was presented. The mean d' obtained from these data is positive and significantly different from zero (mean $d' = 0.49$; $t = 6.746$; $p < 0.001$). In addition to this, this value is close to the empirical threshold of 0.5 indicating, this time that the authenticity responses clearly differed depending on whether the preferred or the less liked product was presented. We can thus conclude that, when measured over the whole panel, the preferred product is actually the most "authentic". This indicates a strong link between the degree of authenticity and the degree of liking for these products.

The authenticity responses from the consumers who preferred the national brand can be compared with those from the consumers who stated that they prefer the private label. As expected, participants who preferred the national brand gave more 'original' responses to the national brand product (48%) than to the private label (29%). In contrast, participants who preferred the private label product gave more 'original' responses to the private label (49%) than to the national brand (34%).

3.3.3 Relationship between authenticity and preference: individual consistency

The link between authenticity and preference could also be investigated by focusing on the subjects who differentiated the two products in terms of authenticity ("discriminating subjects"). Table 2 shows the authenticity/preference consistency for the discriminating subjects belonging to the "national brands identifiers" and "private label identifiers" defined above.

Table 2. Individual authenticity/preference consistency: preference responses (national brand, private label and no preference) for both groups of discriminating subjects.

	Stated preference		
	National brand	Private label	No preference
“National brand identifiers” (71 subjects)	70.4 %	12.7%	16.9%
“Private label identifiers” (85 subjects)	31.8%	57.6%	10.6%

We can see that the degree of consistency is high for the participants who considered the national brand to be the original product, with 70% of these subjects reporting a preference for the national brand. On the contrary, the authenticity/preference link for participants who identified the private label as the ‘original’ product is much weaker since only 57.6 % of the respondents behaved according to our assumption. In fact, 31.8 % of the subjects identified the private label as being the “original” product but preferred the national brand. Given these individual results, it is possible that, for many subjects, similar psychological mechanisms are triggered when they answered the authenticity and the hedonic question. However, some gaps between authenticity and preference answers were found. In fact, even if a great majority of the discriminating subjects performed the authenticity task according to their affective attitude toward the tested products (consistency with preference), some of them answered in an unexpected way. The “inconsistent” results of these discriminating subjects could be interpreted in different ways. First, the formal experiment conditions may have placed these subjects in an analytical mindset in spite of the upsetting story. Second, contrary to our assumption, sensory expectations in terms of authenticity could be quite different from sensory expectations in terms of liking. This hypothesis is supported by the observation of the preference results of the 88 subjects who did not differentiate the two products in terms of authenticity (“liberal subjects” and “conservative subjects”). In fact, among those who reported a preference, there are marginally (Binomial test, $p = 0.075$) more subjects who stated they preferred the national brand (58.5%) than subjects who stated they preferred the private label (41.5%). Only 6.8% of the non-discriminating subjects reported having “no preference”. Thus, these subjects did not report a difference in authenticity between the two products but clearly expressed a preference. Therefore, our results call into question the relevance of using the perceived authenticity degree as a strict alternative to the hedonic measurement.

The study of the authenticity responses of the 27 subjects who declared having “no preference” at the end of the session also provides insightful information. Surprisingly, their authenticity data show a large difference between the degree of authenticity of the two brands, with the national brand being clearly identified as more authentic (mean $d' = 0.59$; $t = 2.61$; $p = 0.015$). These subjects expressed that they did not prefer one brand even if they easily identified the national brand as the “original” product. Thus, again for these subjects the authenticity and liking sensory expectations could be clearly different. However, the choice of the “no preference” response category could also be due to a specific trait of these subjects. In fact, many interviewees who lack self-confidence and do not like to make up their mind usually choose the “neutral” point when they answer an attitude question (McBride & MacFie, 1990). It is likely that the 27 subjects who reported a “no preference” behaved in this way. Their confidence ratings reported during the authenticity task confirm this lack of self-confidence since their mean rating is significantly lower than that of the subjects who declared a preference ($t = 3.704$; $p < 0.001$).

3.4. OVERALL DISCUSSION

This study reveals that the implementation of our authenticity test design when evaluating liking for low affective value food products has many advantages. First, the repeated exposure design provides judgements that are based on several consumptions of the tested products. This probably leads to a better external validity than that of hedonic data collected from first impression only. This can be related to Berlyne’s categorization of hedonic task response behaviours into either specific or diversive exploration (Berlyne, 1960). Specific exploration is not concerned with pleasure but with resolving puzzling stimulation and reducing uncertainty. Thus, instead of real preference or liking data, when they test first impression, market researchers obtain data that are mostly based on curiosity and the desire to learn more about the products. The diversive exploration focuses on the pleasurable aspect of the product and thus for the more durable appreciation of it. This is only possible after specific exploration is completed and the uncertainties are resolved. Thus, contrary to the non-reliable aspect of data collected on first impression, our repeated exposure protocol may lead to more valid hedonic information. In addition to this, we found that the authenticity test induces the subjects to be more involved when they answered the preference question, which probably was decisive in their judgement making process. This is reflected by the significant preference for the national brand over the private label as we also observed in the HUT. The

authenticity task prior to the hedonic question may contribute to the improvement of the respondents' commitment in their responses even when food products with relatively low affective value are tested. On a more pragmatic note, implementation of such an authenticity protocol is much less time consuming and cheaper than a traditional HUT.

Naturally, this method also bears some dangers. First, we observed that many subjects behaved according to what seems like an “over-suspicious” mindset, preventing them from reporting a difference in authenticity. It would be interesting to test the impact of a positive story on the involvement of the subjects. Furthermore, the proximity of the successive exposures may exclude the use of such a design with food products that have persistent sensory properties. Thus, it would be worth investigating whether the authenticity test is feasible with fewer repetitions or including breaks between tasted samples.

Concerning the way we analyzed these authenticity data, we want to emphasize that the analysis of the overall data would not have been sufficient to obtain our conclusions. In fact, the overall data hid essential information because of the diversity of individual response patterns. Indeed, the very low overall mean d' would have led us to conclude that the subjects did not perceive any difference of genuineness between the two crackers, which is not actually the case. We thus promote that this kind of subgroup analysis that takes the individual response patterns into account is crucial and that Cluster Analysis is a good tool to do it. Moreover, this confirms the need to calculate the d' values on individual SDT tables rather than on an overall SDT table. This subgroup analysis is further enriched with confidence data that allows us to sketch various attitudes.

The main result of the present experiment concerns the link between the authenticity and preference data, which is discussed following. At the group level, a link between the preferences and the perceived degrees of authenticity of the products was found since the national brand is the preferred product and is also considered to be the most authentic product. Besides, the over-scoring effect for the warm-up sample illustrates a common feature between the psychological mechanisms triggered to answer an authenticity question and a hedonic question. However, on an individual basis, it seems that the sensory expectations in terms of authenticity can be different from the sensory expectations in terms of liking. Hence, we believe that this kind of paired authenticity/preference data brings some insightful information, providing insight in the diversity of consumer perception. Actually, food developers might be interested in both the liking and the perceived authenticity of their recipes, especially when the results are not correlated. For example, it could be carried out with regular users of the food product in order to evaluate the consequences of a recipe-

change on consumers' sensory expectations, both in terms of authenticity and in terms of liking. In the same way, this approach could be well appropriated when investigating the links between consumer judgement and choice behaviour with variables such as brand or price. Implementation of this authenticity task to test innovative products could also be very informative, by measuring sensory expectations conveyed by the innovative concept. In such a task, consumers would have to guess the authenticity of formulated recipes that they might never have experienced before.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Eleonore Loesher for carefully reading this paper.

REFERENCES

- Berlyne, D. E. (1960). *Conflict, arousal and curiosity*. New York: McGraw-Hill.
- Frandsen, L. W., Dijksterhuis, G., Brockoff, P., Nielsen, J. H., & Martens, M. (2003). Subtle differences in milk: comparison of an analytical and affective test. *Food Quality and Preference*, *14*, 515-526.
- Frandsen, L. W., Dijksterhuis, G., Brockoff, P., Nielsen, J. H., & Martens, M. (2007). Feeling as a basis for discrimination: comparison of a modified authenticity test with the same-different test for slightly different types of milk. *Food Quality and Preference*, *18*, 97-105.
- Köster, E. P. (1998). Les épreuves hédonique. In: F. Depledge & F. Strigler (Eds.), *Evaluation Sensorielle, Manuel Methodologique (2nd ed.)*, (pp. 182-206). Paris: Lavoisier.
- Köster, M. A., Prescott, J., & Köster, E. P. (2004). Incidental learning and memory for three basic tastes in food. *Chemical Senses*, *29*, 441-453.
- Lévy, C. M. & Köster, E. P. (1999). The relevance of initial hedonic judgements in the prediction of subtle food choices. *Food Quality and Preference*, *10*, 185-200.
- Lévy, C. M., MacRae, A. W. & Köster, E. P. (2006). Perceived stimulus complexity and food preference. *Acta Psychologica*, *123*, 394-413.
- McBride, R. L. M. & MacFie, H. J. H. (1990). *Psychological basis of sensory evaluation* (pp. 118-119). London: Elsevier Applied Science.
- Meiselman, H. L., deGraaf, C., & Leshner, L. L. (2000). The effects of variety and monotony on food acceptance and intake at a midday meal. *Physiology & Behavior*, *70*, 119-125.
- Mojet, J. & Köster, E. P. (2005). Sensory memory and food texture. *Food Quality and Preference*, *16*, 251-266.
- Mojet, J., & Köster, E. P. (1986). Onderzoek naar de waardering van drie laag-alcoholische bieren (Research on the appreciation of three low-alcohol beers). Report Psychological Laboratory, Utrecht University.
- Møller, P., Wulff, C. & Köster, E. P. (2004) Do age differences in odour memory depend on differences in verbal memory? *Neuroreport*, *15*, 915-917.
- Stone, H., & Sidel, J. L. (1993). *Sensory Evaluation Practices (2nd ed.)*. San Diego, California: Academic Press Inc.
- Zajonc, R. B. (1980). Feeling and thinking. Preferences need no inferences. *American Psychologist*, *35*, 151-175.

II - LES METHODES NON-HYPOTHETIQUES

Après avoir présenté le principe général des méthodes non-hypothétiques, nous détaillerons les différents types de méthodes qui existent. Nous expliquerons ensuite le type de protocole que nous avons choisi de privilégier dans ce travail pour révéler la préférence des consommateurs lors d'un test en salle. Enfin, l'application de ce protocole dans le cadre d'un test à grande échelle sera présentée sous forme de publication.

II.1 - PRINCIPE ET DESCRIPTION DES METHODES NON-HYPOTHETIQUES

II.1.1 - Principe général

Comme nous l'avons précisé précédemment, nous remettons en question la qualité des mesures hédoniques directes basées sur la déclaration d'une opinion. En effet, un individu ne sait pas toujours ce qui guide ses choix et beaucoup de ses décisions pourraient même être contraires à ce qu'il pense qu'il devrait faire (Shiv et Fedorikhin, 1999). Ainsi, même si les méthodes fondées sur l'expression des préférences sont souvent plus faciles à appliquer, il convient d'utiliser des mesures indirectes pour révéler des préférences. Décrire et expliquer le comportement des consommateurs est un domaine d'étude que nous partageons avec les économistes qui ont, de leur côté, développé des méthodes permettant de révéler les préférences. Plus précisément, les économistes s'intéressent à la valeur allouée à un produit par les consommateurs. Ainsi, une des méthodes d'investigation pour étudier le potentiel d'un produit sur le marché consiste à estimer les valeurs économiques privées des consommateurs (Hicks, 1945) comme le consentement à payer (*Willingness To Pay*). Il s'agit en fait d'accéder au prix de réserve du consommateur, c'est-à-dire le prix maximum qu'il est prêt à payer pour obtenir un bien ou un service. Cette mesure du consentement à payer peut être obtenue par le biais notamment des méthodes d'analyse conjointe. Cependant, ces méthodes sont généralement mises en place avec des produits hypothétiques c'est-à-dire avec des scénarios décrivant les attributs des produits lors de problématiques de fixation de prix pour des produits nouveaux ou d'enquêtes concernant l'impact d'informations (utilisation d'OGM, pays d'origine) sur le comportement alimentaire (Alfnes, 2004 ; Burton *et al.*, 2001 ; Lusk *et al.*, 2003). Les consentements à payer obtenus dans ces études ne sont pas basés sur un achat réel des produits et restent donc des mesures déclaratives au même titre que l'appréciation ou la préférence déclarée. En effet, des consommateurs peuvent déclarer des valeurs de

consentement à payer élevés pour des produits ayant une qualité perçue élevée sans pour autant les acheter réellement une fois qu'ils sont placés face aux produits et qu'ils doivent dépenser leur argent. Les économistes parlent alors de mesures hypothétiques en opposition aux méthodes de mesure non-hypothétiques qu'ils ont développé dont la réponse collectée induit justement des conséquences cognitives et monétaires pour le participant. La littérature florissante dans le domaine de l'économie expérimentale démontre la nécessité de travailler avec des méthodes non-hypothétiques pour révéler des « vrais » comportements. En effet, de nombreux auteurs ont mis en évidence que les consentements à payer hypothétiques pour des produits sont plus élevés que des consentements à payer non-hypothétiques (Fox *et al.*, 1998 ; List et Shogren, 1998 ; Lusk, 2003 ; Lusk et Schroeder, 2004). Lors d'études hypothétiques où l'on demande aux sujets de déclarer leur prix de réserve, les participants peuvent essayer de deviner l'objectif de l'étude ainsi que les conséquences de leurs réponses ce qui peut entraîner une modification de leur « vrai » prix de réserve. Si le sujet pense que sa réponse sera utilisée pour fixer le prix de vente d'un produit, il aura tendance à sous-évaluer son prix de réserve. S'il pense que l'étude aura pour conséquence le lancement sur le marché d'un nouveau produit qu'il désire ardemment, il peut sur-évaluer son prix de réserve. Les méthodes non-hypothétiques permettent de révéler les valeurs économiques privées et donc les préférences des individus sur la base de l'achat de vrais produits. Les participants sont rémunérés et peuvent utiliser cet argent pour acheter des produits. Même si de nombreuses procédures ont été développées dans le domaine de l'économie expérimentale pour accéder à ces prix de réserve non-hypothétiques, elles présentent deux caractéristiques communes.

- Elles introduisent du réalisme dans les tests en mettant en jeu de vrais produits et de l'argent et en demandant au participant de réaliser un vrai compromis entre un bien et de l'argent.
- Elles proposent des mécanismes qui font que le participant, s'il veut maximiser son profit, a tout intérêt à déclarer précisément son « vrai » prix de réserve (c'est-à-dire ses « vraies » préférences). Les économistes parlent de règles incitatives car :
 - la sous-évaluation du prix de réserve peut faire passer le participant à côté d'un achat particulièrement intéressant (une affaire),
 - et la sur-évaluation du prix de réserve peut avoir la conséquence d'obliger le participant à acheter un produit à un prix trop élevé pour lui.

Dans notre objectif de propositions méthodologiques visant à accéder de manière plus fidèle aux préférences des consommateurs lors d'un test en salle, ces méthodes basées sur le principe de l'engagement des répondants, nous ont paru tout à fait prometteuses.

II.1.2 - Les différents types de méthodes non-hypothétiques

L'examen de la littérature nous a permis d'identifier deux grands types de méthodes non-hypothétiques : les enchères expérimentales et les choix expérimentaux non-hypothétiques. Ces deux méthodes se différencient principalement par le fait que les premières permettent de récolter directement le prix de réserve du sujet alors que les deuxièmes déduisent généralement les prix de réserve en fonction de choix faits sous contrainte économique.

II.1.2.1 - Les enchères expérimentales

L'application des enchères expérimentales dans le domaine de l'agroalimentaire pour évaluer les valeurs économiques privées des consommateurs est fréquente (Buzby *et al.*, 1998 ; Fox, 1995 ; Hayes *et al.*, 1995 ; Lusk *et al.*, 2001 ; Melton *et al.*, 1996). Comme leur nom l'indique, les enchères expérimentales ont pour objectif de demander directement aux participants de faire une offre. Il existe différentes situations d'enchères qui permettent d'aboutir à différents types de mesures. Le tableau 19 présente les différentes variantes de protocoles et les mesures auxquelles elles aboutissent ainsi que quelques études publiées dans *Food Quality and Preference* les ayant mis en pratique.

Tableau 19 : Différents protocoles de mesure pour révéler les valeurs économiques privées

		DOTATION INITIALE	
		NON	OUI
WTP le participant agit comme un acheteur	Prix maximal que le sujet est prêt à payer pour acheter un produit (Noussair <i>et al.</i> , 2004a; Lange <i>et al.</i> , 2002; Stefani <i>et al.</i> , 2006)		Prix maximal que le sujet est prêt à payer pour échanger un produit qu'il possède pour un produit de meilleure qualité.
WTA le participant agit comme un vendeur	<i>Non utilisé</i>		Compensation minimale que le sujet souhaite recevoir pour accepter d'échanger un produit qu'il possède pour un produit de moins bonne qualité (Jaeger <i>et al.</i> , 2004; Lusk <i>et al.</i> 2001)

Deux types d'offres peuvent donc être demandées au sujet : un prix d'achat (*willingness to pay* : *WTP*) ou un prix de vente (*willingness to accept* : *WTA*). Une offre correspondant à un prix de vente nécessite la dotation d'un bien au sujet avant qu'il fasse son offre. En revanche, une offre d'achat peut être réalisée dans le cadre de la dotation d'un produit ou non. Détaillons ces différentes variantes de protocoles et leur utilité.

Dotation d'un produit ou non

L'examen de la littérature sur les méthodes d'enchères expérimentales a mis en évidence une caractéristique importante du protocole qui est de donner ou non un produit au sujet avant la tâche d'enchère. En effet, même si certaines études ont pour objectif d'accéder au consentement à payer des consommateurs pour un produit dans l'absolu, de nombreuses études souhaitent comparer les valeurs allouées à différents objets ou au même objet entouré de différentes informations. Dans un tel objectif de comparaison, l'expérimentateur a deux possibilités :

- soit il met en place plusieurs mécanismes d'enchères sans dotation et révèle le consentement à payer des sujets pour chaque produit qu'il compare par la suite,
- soit il essaye de récupérer directement cette différence de prix de réserve.

En fait, de nombreuses études s'attachent à révéler directement la valeur allouée à l'information ou à un produit en référence à celle d'un autre produit (Hoffman *et al.*, 1993). Pour cela, le sujet est doté d'un produit et l'expérimentateur propose au sujet d'offrir un prix pour échanger son produit pour un produit de qualité supérieure.

L'avantage de la dotation est qu'elle force le participant à être plus attentif aux offres qu'il fait (Fox, 1995). Lusk *et al.* (2001) précisent également que l'approche avec dotation est particulièrement attractive pour le recrutement des sujets. En dépit de ces avantages, l'approche avec dotation peut introduire un biais dans l'estimation du consentement à payer. En effet, Tversky et Kahneman (1991) précisent que les individus peuvent allouer plus de valeur à un bien s'ils le possèdent que s'ils ne le possèdent pas. Cet effet de dotation provient de la tendance à l'aversion pour les pertes c'est-à-dire que des pertes sont généralement évaluées comme plus élevées que des gains d'un montant équivalent. Lusk *et al.* (2004) ont mesuré cet effet de dotation en comparant la différence des consentements à payer de deux produits A et B obtenus sans dotation ($WTP_B - WTP_A$) avec le consentement à payer pour échanger le produit A contre le produit B (WTP_{AB}). Ils observent alors que $WTP_{AB} < WTP_B - WTP_A$, ce qui signifie que les sujets évaluent le produit A plus fortement quand il est en leur possession que quand ils doivent faire une offre pour l'acquérir.

Consentement à payer ou consentement à accepter (WTP ou WTA)

Depuis le début de ce paragraphe sur la mesure des valeurs économiques, nous évoquons principalement la mesure du consentement à payer (WTP). Cependant une autre valeur économique privée peut être mesurée lors de ces marchés expérimentaux dans le cas d'une approche avec dotation. Il s'agit du consentement à accepter (WTA) qui correspond à la

compensation minimum qu'un sujet souhaite recevoir pour accepter d'échanger un produit qu'il possède contre un autre de moins bonne qualité. Dans ce type d'enchère le sujet n'agit alors plus comme un acheteur mais comme un vendeur. Même si cette mesure du consentement à accepter est beaucoup moins utilisée que celle du consentement à payer, elle est particulièrement adéquate pour évaluer l'attitude des consommateurs face à des informations négatives comme les aliments avec des OGM.

La comparaison des valeurs de consentement à payer et à accepter sur les mêmes produits a fait l'objet de nombreux travaux dont les résultats troublent les économistes. En effet, Horowitz et McConnell (2002) observent sur un ensemble de 45 études une forte divergence entre les valeurs de WTP et de WTA qui est particulièrement dérangeante puisqu'elle remettrait en question la validité et le caractère incitatif des méthodes d'enchères pour révéler les « vraies » valeurs de réserve (Knetsch *et al.*, 1998). Tversky *et al.* (1990) rejettent cependant l'idée que cette divergence provient d'un artefact des mécanismes pour révéler ces valeurs et proposent de l'expliquer de la même façon que l'effet de dotation c'est-à-dire par la tendance d'aversion pour les pertes.

L'enchère BDM (Becker et al., 1964) : un mécanisme incitatif individuel

Il existe de nombreux mécanismes pour inciter le sujet à révéler son « vrai » prix de réserve comme les enchères au second prix de Vickrey (1961) ou les enchères au ⁿ^{ième} prix aléatoire de Shogren *et al.* (2001). Contrairement à ces deux méthodes qui s'appliquent dans le cadre d'un groupe de participants, la méthode BDM (Becker *et al.*, 1964) propose des enchères individuelles. Cette caractéristique nous a particulièrement intéressé dans le cadre de notre travail étant donné que les séances en groupe sont beaucoup plus coûteuses et difficiles à mettre en place. La méthode BDM n'est pas à strictement parler une méthode d'enchères mais elle présente les mêmes règles incitatives que les enchères précédentes basées sur le fait que le prix de vente du produit ne dépend pas du prix indiqué par le sujet.

La figure 25 présente le déroulement d'une enchère BDM dans le cas de la mesure d'un consentement à payer sans dotation. Le participant offre un prix d'achat qui correspond au prix maximum qu'il est prêt à payer pour acquérir le bien. Il tire ensuite un prix de vente dans une distribution de prix allant de zéro à un prix supérieur au prix maximum anticipé que pourrait offrir un sujet. Si son prix d'achat est supérieur ou égal à ce prix de vente tiré au sort, le participant achète le produit au prix de vente et fait donc une bonne affaire puisqu'il était prêt à payer plus cher. Si son prix d'achat est inférieur au prix de vente, le participant ne peut pas acheter le produit mais il ne regrette pas puisque ce prix de vente est trop élevé pour lui.

Cette méthode ne nécessite pas de séance d'entraînement stricte pour comprendre les règles d'incitation. Il suffit juste de laisser l'opportunité au sujet de modifier son offre en lui proposant d'imaginer qu'il tire un prix de vente juste au-dessus de son offre. Il comprend alors qu'il ne pourrait pas acheter le produit à ce prix de vente. Si jamais il montre un sentiment de regret de ne pas pouvoir acheter le produit à ce prix de vente, cela signifie que sa première offre était sous-estimée et on lui laisse l'opportunité de la modifier.

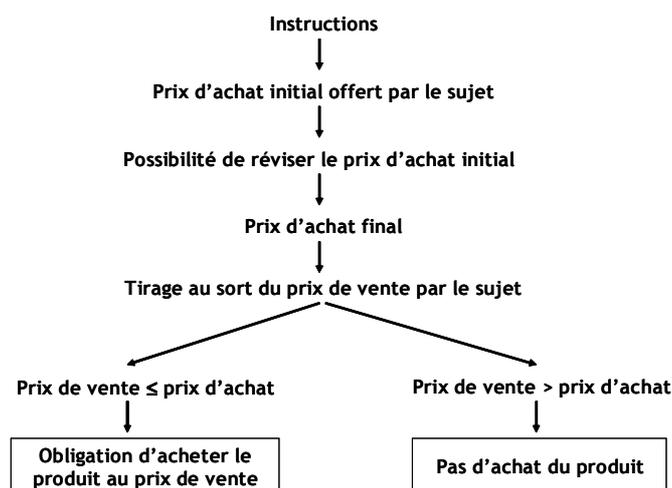


Figure 25 : Démarche globale de la méthode BDM

Cette méthode a été employée en laboratoire (Noussair *et al.*, 2004b ; Nunes et Boatwright, 2004) mais sa simplicité en a fait également son succès pour évaluer les consentements à payer de consommateurs sur des lieux de vente (Lusk *et al.*, 2001 ; Lusk et Fox, 2003 ; Wertenbroch et Skiera, 2002). Sur un lieu de vente, les participants sont déjà dans un état d'esprit d'achat et payent les produits avec leur argent ce qui augmente d'autant plus leur engagement. En effet, lors des deux autres méthodes d'enchères (Vickrey et n^{ième} prix aléatoire) les participants pré-recrutés reçoivent une rémunération et ne quittent donc jamais la séance avec moins d'argent qu'à leur arrivée.

II.1.2.2 - Les choix expérimentaux non-hypothétiques

Les choix expérimentaux non-hypothétiques correspondent à des choix avec une motivation économique réaliste. Les participants doivent faire des choix dans une série de scénarios proposant plusieurs alternatives de produits étiquetés différemment. La littérature met en évidence deux types de compromis auxquels le participant est confronté lors de ces choix non-hypothétiques : un compromis qualité/prix ou un compromis qualité/quantité.

Un compromis qualité/prix

Dans chaque scénario, deux produits ou plus sont présentés avec des propositions de prix de vente différents. Une fois que le participant a fait son choix pour chaque scénario, il en tire un au sort. Le produit qu'il a choisi dans ce scénario est donc acheté au prix indiqué. Le mécanisme incitatif de ce protocole réside dans le tirage aléatoire du scénario effectif. Les participants doivent évaluer de manière très attentive chacun des scénarios étant donné qu'ils ont tous autant de chance d'être tirés au sort.

Ces méthodes de choix expérimentaux sont généralement utilisées pour étudier l'impact de certaines caractéristiques des aliments sur les consentements à payer des consommateurs. Dans ce cas-là, les alternatives de choix présentent le même produit qui est proposé à un prix différent en fonction d'une caractéristique différente. De cette façon, Lusk et Schroeder (2004) ont par exemple étudié l'impact du label de qualité sur le consentement à payer pour un steak et Alfnes *et al.* (2007) ont étudié l'impact de la couleur de saumon proposée sur le consentement à payer des consommateurs. Dans le cas d'une expérience de choix qui propose réellement différentes alternatives de produits à des prix différents, les participants doivent donc faire un compromis entre de l'argent et la qualité d'un produit. Le principal avantage de ces méthodes de choix, comparé aux enchères, est qu'elles proposent une procédure relativement simple à comprendre par les consommateurs qui sont quotidiennement confrontés à ce genre de décision dans un magasin proposant différentes alternatives à différents coûts.

Un compromis qualité/quantité

Il existe également des études basées sur un processus d'échange de produits pour révéler la valeur d'un bien par rapport à celle d'un autre (Marette *et al.*, 2006 ; Sanogo et Masters, 2002). Plutôt que d'échanger de l'argent contre un produit, il s'agit d'échanger des lots de produits en quantités différentes. Les participants doivent réaliser un compromis entre la quantité et la qualité de produits à emporter à la maison.

Sanogo et Masters (2002) ont utilisé ce design basé sur des échanges de produit pour évaluer, entre autres, l'impact de la marque sur les consentements à payer de femmes maliennes pour un aliment infantile. Au départ, chaque participante reçoit 400 g du produit de marque. Dans la première alternative de choix, la participante a la possibilité d'échanger ses 400 g de produit de marque pour 400 g du produit sans marque. Si jamais celle-ci n'accepte pas l'échange, l'alternative suivante lui est proposée en augmentant l'offre du produit sans marque de 100 g. La quantité du produit sans marque est ainsi incrémentée jusqu'à ce que la

participante accepte l'offre (avec un maximum à 900 g). Les choix aux différentes alternatives sont réalisés sur des questionnaires séparés qui sont déposés dans une urne. A la fin du test, la participante tire au sort un des scénarios. Ainsi, chaque choix a autant de chance de déterminer ce que la participante ramènera à la maison : 400 g du produit de marque ou un lot du produit sans marque d'une valeur similaire.

L'idée développée par Sanogo et Masters (2002) est d'utiliser les choix des participants pour révéler leur consentement à payer pour le produit sans marque sur la base de la quantité nécessaire pour induire un échange (c'est-à-dire l'indifférence révélée par le consentement à échanger des participants). Marette *et al.* (2006) ont utilisé le même type de protocole pour évaluer l'impact d'informations sur le choix de thon ou de sardines en boîte. La particularité de cette étude est que les expérimentateurs ne se basent pas sur une préférence a priori d'un produit sur l'autre. Ainsi, la moitié des sujets reçoit six boîtes de thon et l'autre moitié des sujets reçoit six boîtes de sardines. Les sujets dégustent ensuite les deux produits et doivent choisir de conserver leurs six boîtes ou de les échanger contre une à douze boîtes de l'autre produit lors de différents scénarios.

II. 2 - CHOIX DU PROTOCOLE NON-HYPOTHETIQUE A PRIVILEGIER

De la même façon que pour la méthode du test d'authenticité, nous souhaitons que le protocole non-hypothétique sélectionné réponde aux critères suivants :

- Des contraintes budgétaires identiques à un test en salle classique.
- Un protocole facile à mettre en œuvre par des enquêteurs non expérimentés.
- Une tâche facile à comprendre et à réaliser par les participants en peu de temps.
- Un protocole qui fonctionne avec des produits peu impliquants (à faible valeur monétaire).

Les contraintes budgétaires nécessitent des interviews individuelles de vingt minutes maximum afin de conserver un recrutement en rue beaucoup moins coûteux qu'un pré-recrutement. L'enchère BDM ainsi que les choix expérimentaux basés sur un compromis prix/qualité seraient envisageables puisque ces deux méthodes permettent une interview en face-à-face. Cependant les mécanismes incitatifs de ces méthodes restent basés sur des conséquences monétaires et peuvent sembler difficiles à appliquer dans le cas de participants recrutés ponctuellement pour tester des produits peu impliquants comme des crackers salés.

Premièrement, ces méthodes sont, pour la majorité, utilisées en laboratoire après un pré-recrutement des sujets qui arrivent donc avec la ferme intention d'acheter un produit. Dans le cas d'un recrutement en rue, les participants ne sont pas préparés à un tel achat et on peut donc s'interroger sur leur réelle motivation à dépenser l'argent qu'on leur donne en début d'interview pour l'achat de produits alimentaires à faible valeur unitaire (même s'il s'agit d'une affaire). Même si les enchères BDM ont déjà été appliquées pour évaluer des produits à faible valeur monétaire comme des barres chocolatées ou des canettes de soda (Lusk *et al.*, 2001 ; Lusk et Fox, 2003 ; Wertenbroch et Skiera, 2002), le recrutement ponctuel des participants a été systématiquement réalisé sur des lieux d'achat du produit testé, garantissant ainsi un réel désir des participants de repartir avec le produit en question et diminuant donc le risque que les participants sous-évaluent leur offre pour ne pas dépenser d'argent.

Deuxièmement, ces méthodes non-hypothétiques basées sur des conséquences monétaires posent des problèmes logistiques et budgétaires. Elles nécessitent une rémunération des participants ce qui augmente le coût total de l'étude. Par ailleurs, la rémunération en argent liquide pose un problème aux instituts d'études qui ne sont pas organisés pour l'introduction de monnaie dans une salle de test. Enfin, les instituts ainsi que l'entreprise qui finance l'étude peuvent être réticents à vendre des produits concurrents. Ainsi, même si ces méthodes non-hypothétiques introduisant une conséquence monétaire pour les participants sont reconnues pour être efficaces lors de problématiques de recherche ponctuelles, elles n'ont pas reçu un accueil très enthousiaste de la part des industriels et des instituts d'études. Lange *et al.* (2002) reconnaissent que les enchères expérimentales demandant aux sujets d'acheter des produits sont particulièrement sensibles pour mettre en évidence des différences de valeurs perçues entre des produits se différenciant sur des aspects extrinsèques comme la marque ou les informations externes. Cependant, très peu de travaux ont utilisé ces méthodes d'enchères dans le cadre d'évaluation de produits en aveugle sans y associer aucune information externe. L'introduction d'argent dans un test de produit sans information externe pourrait entraîner les sujets à se désintéresser de la recette des produits pour se focaliser plus sur des aspects liés à la possibilité de « faire une affaire ». Ce comportement serait d'autant plus probable dans le cas où il n'y aurait pas un rejet très clair de l'une des alternatives (rejet dû à la recette) ce qui est très souvent le cas lors d'un test hédonique en aveugle. Ceci pourrait donc remettre en cause le pouvoir discriminant des valeurs de réserves récoltées par ces méthodes d'enchères dans le cas de produits plutôt proches sensoriellement.

La méthode des choix expérimentaux non-hypothétiques dont l'implication des participants est basée sur la conséquence cognitive liée à la quantité de produit que le participant peut ramener chez lui pourrait être plus efficace pour conserver l'attention du sujet sur les produits dégustés sans information externe. Cette méthode a l'avantage de permettre une certaine implication des participants dans leur réponse sans nécessiter l'introduction d'argent lors des tests. L'implication ici est d'autant plus forte que les conséquences cognitives sont liées aux produits testés et non à de l'argent gagné ou perdu.

Nous avons donc choisi de mettre en place un tel protocole proposant au participant d'échanger différentes quantités de deux produits afin d'évaluer une préférence entre deux produits et l'intensité de cette préférence. Au lieu d'évaluer les « prix de réserve », ce protocole nous permettra d'accéder aux « quantités de réserve » relatives des participants.

II.3 – APPLICATION DU PROTOCOLE POUR COMPARER LES PERFORMANCES ORGANOLEPTIQUES DE DEUX BISCUITS FEUILLETES AU FROMAGE (ARTICLE 4)

Le développement du protocole basé sur le principe d'un échange de quantités différentes des deux produits a nécessité la réalisation d'une phase de pré-tests avec 50 sujets. Cette phase de pré-tests a été notamment l'occasion de vérifier sa faisabilité technique par les participants et par des enquêteurs non expérimentés en vingt minutes d'interview. Par ailleurs, de nombreux aspects méthodologiques ont pu également être mis en place :

- le discours explicatif à employer (type de vocabulaire),
- le format des questionnaires de choix (illustrations des quantités ou non),
- les quantités de produit à mettre en jeu (les quantités initiales, le pas des différentes offres, les bornes minimales et maximales),
- le déroulement de l'épreuve (ordre de présentation des différentes offres).

L'article présenté ci-après présente l'application de ce protocole d'échange de produits pour comparer les performances organoleptiques des deux biscuits feuilletés au fromage.

ARTICLE 4 : Boutrolle I., Combris P., Abiven F., Arranz D., Delarue J., Danzart, M.

Proposition d'amélioration des protocoles de test de produits alimentaires.

Revue Française de Marketing, accepté en octobre 2007.

RESUME

L'objectif de cette étude est de proposer un protocole de test amélioré pour révéler les préférences des consommateurs envers deux recettes de biscuits salés. Le principe de ce protocole repose sur une tâche de choix de produits à emporter à la maison. Après avoir dégusté les deux biscuits, un premier choix entre 300 g de chaque biscuit est proposé au participant. Différentes propositions de choix construites de manière à ce que la quantité du biscuit préféré soit en compétition avec une quantité supérieure du biscuit moins aimé sont ensuite faites au participant. Bien que la question de préférence posée lors d'un test de produit traditionnel ne montre aucune préférence significative envers l'une des deux recettes de biscuits, les résultats de ce test de choix montrent au contraire qu'un produit est significativement plus souvent choisi (60.2%) que l'autre (39.8%). Par ailleurs, les variations minimales de quantité nécessaires pour entraîner un échange du biscuit préféré pour le biscuit moins aimé sont utilisées pour quantifier l'intensité de la préférence des consommateurs pour chaque produit.

Mots clés : *Préférence, Choix, Consentement à Echanger, Implication, Produits alimentaires*

ABSTRACT

The aim of this study was to propose an improved test protocol to measure consumers' preferences for low added-value food products, namely two recipes of salted crackers. The basic idea was to set up a non-hypothetical choice experiment. After having tasted the two crackers, the participant was presented with a first choice between 300g of each product. Afterwards, sequential alternative propositions requesting a choice between different amounts of each product were presented. Each proposition presented a quantity of the preferred product in competition with a higher quantity of the less liked product. Although the hypothetical preference question asked in a traditional preference test did not show any preference, our results show that one product is significantly more frequently chosen (60.2%) than the other one (39.8%). In addition, the minimum variations of quantity needed to induce a swap of the preferred product for the less liked product are used to quantify the intensity of the preference for each product.

Keywords: *Preference, Choice, Willingness to Swap, Involvement, Food Products*

**PROPOSITION D'AMELIORATION DES PROTOCOLES DE TEST DE PRODUITS
ALIMENTAIRES.**

Isabelle Boutrolle^{a,d}, Pierre Combris^b, François Abiven^c, Delphine Arranz^a, Marc Danzart^d,
Julien Delarue^d.

^a*Consumer Science - Danone Research, RD 128, 91767 Palaiseau, France.*

^b*Laboratoire de Recherche sur la Consommation, INRA, 65 Boulevard de Brandebourg, 94205 Ivry-sur-Seine
Cedex, France*

^c*Repères, Institut d'études marketing - 9 rue Rougemont, 75009 Paris, France*

^d*Laboratoire de Perception Sensorielle et Sensométrie, AgroParisTech (site de Massy), 1 avenue des
Olympiades, 91744 Massy, France.*

INTRODUCTION

Les entreprises agroalimentaires soucieuses d'améliorer leur position concurrentielle souhaitent évaluer les performances de leurs produits par rapport à celles des produits concurrents. Les données du marché (données de vente) sont de bons indicateurs du succès ou de l'échec d'une marque par rapport à une autre. Ces données sont généralement analysées par le marketing qui, dans le cas de mauvais résultats, souhaite tout d'abord vérifier si cet échec ne provient pas d'une mauvaise performance organoleptique de la recette du produit en comparaison à celle du concurrent. Pour cela, les entreprises font appel à des instituts d'études qui mettent en place des tests consommateurs en aveugle dits « tests hédoniques ». Il s'agit de tests gustatifs le plus souvent réalisés sans référence au prix ou à la marque de l'aliment dégusté. La méthodologie mise en place lors de ces tests hédoniques consiste à sélectionner dans des rues commerçantes des consommateurs de la catégorie de produit testée et à les inviter à participer immédiatement à une séance de dégustation dans une salle de test. Les participants dégustent alors de manière séquentielle les différentes recettes et doivent indiquer celle qu'ils préfèrent ou encore, selon les cas, donner une note d'appréciation globale pour chaque recette évaluée. La qualité de ces données quantitatives est cruciale étant donné que les résultats de ces tests hédoniques sont très souvent utilisés pour mettre en place des décisions stratégiques concernant le produit.

CONTEXTE : REMISE EN QUESTION DE LA VALEUR PREDICTIVE DES MESURES HEDONIQUES

L'utilisation de ces tests hédoniques comme outils d'aide à la décision pour les entreprises agro-alimentaires suppose que la mesure de la préférence déclarée entre deux produits prédise un comportement futur de choix des consommateurs (toutes caractéristiques extrinsèques aux recettes identiques par ailleurs). Cependant, la littérature regorge d'études dénonçant la faible capacité prédictive de ces mesures hédoniques déclaratives concernant le comportement des consommateurs. Dès les années 60, quelques études (Schutz, 1957 ; Peryam *et al.*, 1960) pointaient déjà du doigt le fait que les évaluations hédoniques déclaratives seules prédisent au mieux 50% de la variabilité de la consommation. Plus récemment, Levy et Köster (1999) et Leon *et al.* (1999) ont observé des incohérences entre ces données hédoniques déclaratives et le choix réel entre différentes alternatives de produits. De la même façon, Kalwani et Silk (1982) ont également mis en évidence l'échec de la mesure d'intention d'achat pour prédire un achat réel. Ainsi, la faible validité externe de ces données déclaratives entraînant de mauvaises décisions stratégiques serait une des causes de nombreux échecs de produits sur le marché. Un des principaux écueils de ces techniques de collecte de données de préférence est l'absence d'engagement des interviewés lorsqu'ils déclarent aimer ou préférer un produit. En effet, contrairement à un véritable acte d'achat ou de consommation d'un aliment, la réponse donnée n'a pas de conséquence économique ou cognitive pour le participant. Le choix lors de l'acte d'achat du produit implique des affects de satisfaction et de regret pour le consommateur qui n'interviennent pas au niveau de la déclaration d'une simple note d'appréciation ou d'une préférence.

LE BIAIS DE « SATISFICING » DANS LES ETUDES CONSOMMATEURS

Outre le manque de conséquences liées aux réponses données par les participants, les conditions de test, souvent favorables à la distraction (un lieu agité, pas de rémunération et très peu de temps pour répondre au questionnaire), contribuent également au manque d'implication des participants dans la tâche qui leur est demandée. L'ensemble de ces conditions favorise la manifestation du biais dit de « satisficing » (Simon, 1956) qui correspond à la tendance à mettre en œuvre un processus de traitement des informations moins rigoureux et efficace qu'il ne l'aurait été dans des conditions réalistes. Cette stratégie permet à l'individu d'aboutir rapidement à un résultat qui est assez bon mais qui n'est pas forcément optimal (c'est-à-dire le plus satisfaisant). Ainsi, la mauvaise validité externe des données déclaratives proviendrait entre autre de ce biais de « satisficing » se manifestant dans les réponses données dans les enquêtes par :

- le choix de l’option explicitement offerte de réponse « sans opinion »,
- le choix des réponses socialement souhaitables,
- le manque de discrimination entre de multiples objets d’intérêts,
- le biais d’acquiescement qui est la tendance à être d’accord avec n’importe quelle affirmation indépendamment de son contenu.

COMMENT IMPLIQUER LE PARTICIPANT DANS SA REPONSE : LES METHODES

EXPERIMENTALES NON-HYPOTHETIQUES

Décrire et expliquer le comportement des consommateurs est un domaine d’étude exploré également par les économistes qui ont de leur côté développé des méthodes permettant de révéler les préférences. Plus précisément, les économistes s’intéressent à la valeur allouée à un produit par les consommateurs. Ainsi, une des méthodes d’investigation pour étudier le potentiel d’un produit sur le marché consiste à estimer les valeurs économiques privées des consommateurs et en particulier leurs prix de réserve, c’est-à-dire le prix maximal que le consommateur est prêt à payer pour acquérir un bien (Consentement à Payer) ou le prix minimum auquel il est prêt à céder un bien (Consentement à Recevoir ou à Accepter)⁷. Cependant, nous retrouvons dans la littérature florissante de l’économie expérimentale le problème de la faible validité prédictive de ces valeurs de prix de réserve déclarées (Fox et al, 1998 ; List et Shogren, 1998 ; Lusk et Shroeder, 2004 ; Lusk, 2003). Les participants peuvent en effet déclarer des valeurs de WTP élevées pour des produits sans pour autant les acheter réellement une fois qu’ils doivent dépenser leur argent. Les économistes proposent donc d’utiliser des méthodes expérimentales dites « non-hypothétiques » pour révéler ces prix de réserve (Hoffman *et al.*, 1993). Ces méthodes proposent au participant de réaliser un achat réel de produit. Cet acte d’achat a des conséquences monétaires et cognitives pour le participant qui permettent donc d’augmenter son implication dans sa réponse. Les prix de réserve révélés par ces méthodes expérimentales sont supposés être de bons indicateurs des véritables préférences des consommateurs. Ces méthodologies reposent sur des mécanismes incitatifs qui font que si le participant veut maximiser sa satisfaction, sa meilleure stratégie consiste à déclarer précisément son véritable prix de réserve. Les enchères au second prix (Vickrey, 1961) et la méthode BDM (Becker *et al.*, 1964) sont les deux mécanismes incitatifs les plus couramment utilisés. L’enchère au second prix doit être mise en œuvre avec un groupe de sujets. Chaque participant soumet une enchère écrite. Le participant qui a soumis

⁷ Dans la suite du texte, les termes et les abréviations de la littérature internationale seront utilisés : Willingness To Pay (WTP) pour le consentement à payer et Willingness To Accept (WTA) pour le consentement à recevoir.

l'enchère la plus élevée acquiert le bien mis en vente, et paye un prix égal à la deuxième enchère la plus élevée. La procédure BDM consiste à inviter chaque participant à indiquer son prix d'achat maximum. Le prix du produit mis en vente est ensuite tiré au sort, tous les participants qui ont indiqué un prix d'achat supérieur ou égal au prix de vente achètent effectivement le produit, et bénéficient de la différence entre les deux prix. Quelle que soit la méthode, le fait que le prix de vente du produit ne dépende pas du prix indiqué par le sujet permet de l'inciter à faire une offre sincère. En effet, la sous-évaluation du prix de réserve risque de faire passer le participant à côté d'un achat particulièrement intéressant et la sur-évaluation du prix de réserve peut avoir la conséquence de l'obliger à acheter un produit à un prix trop élevé pour lui. Ainsi, le participant doit comprendre que la meilleure stratégie qu'il puisse mettre en place lors de ces marchés expérimentaux est de révéler son véritable prix de réserve.

Cependant, la mise en place de ces méthodes de test par les industriels se heurte à de nombreuses contraintes logistiques et budgétaires. En effet, elles nécessitent le paiement des participants qui contribue à l'augmentation du coût de l'étude. Par ailleurs, les instituts d'études sont parfois réticents à l'application de telles méthodes qui nécessitent la vente de produits concurrents à l'entreprise commanditaire de l'étude. Enfin, même si ces méthodologies ont déjà été employées pour révéler des préférences envers des denrées alimentaires comme du champagne (Lange *et al.*, 2002), des fruits génétiquement modifiés (Jaeger *et al.* 2004) ou des produits exotiques (Stefani *et al.*, 2006), ces méthodes ont peu de chance de fonctionner avec des produits de très faible valeur monétaire et affective.

OBJECTIF DE L'ETUDE

L'objectif de cet article est de présenter un protocole de test consommateurs non-hypothétique qui pourrait être utilisé par les instituts d'études pour évaluer les préférences des consommateurs concernant des aliments de consommation courante. Pour cela, le protocole expérimental proposé doit :

- fonctionner avec des produits de faible valeur monétaire,
- être facile à administrer en 20 minutes d'interview par des enquêteurs non expérimentés,
- pouvoir être mis en place avec le même budget qu'un test hédonique traditionnel.

STRATEGIE MISE EN ŒUVRE

Quelques études (Sanogo et Masters, 2002 ; Marette *et al.*, 2006) ont eu recours à des échanges directs pour révéler la valeur d'un bien par rapport à celle d'un autre. Ainsi, plutôt que d'acheter un produit, il s'agit d'échanger des lots de produits en quantités différentes. Les participants doivent réaliser un compromis entre la quantité et la qualité des produits à emporter chez eux. Cette méthode a le principal avantage de ne pas nécessiter l'introduction d'argent lors des tests tout en favorisant l'implication des participants dans leurs réponses. L'implication par rapport aux produits est d'autant plus forte que les conséquences cognitives sont liées aux produits testés et non à de l'argent gagné ou perdu. Nous proposons donc de mettre en place un protocole de test incitatif dans lequel les participants choisissent entre différentes quantités de deux produits révélant ainsi leur « quantité de réserve » plutôt que leur prix de réserve par l'intermédiaire de consentements à échanger.

Cet article présente l'application de cette méthode lors d'un test consommateur pour évaluer la préférence entre deux recettes de biscuits d'apéritif. La préférence révélée par ce protocole est comparée avec les informations hédoniques récoltées lors d'un test hédonique traditionnel.

METHODOLOGIE

LES ECHANTILLONS

Deux études consommateurs ont été menées selon deux protocoles de test distincts. Les participants sont recrutés dans les rues de trois villes Française (Paris, Lyon et Marseille) entre 10 h et 18 h et sont invités à participer à une séance de dégustation de 20 minutes. Les participants recrutés sont des femmes (de 18 à 65 ans) consommatrices de biscuits d'apéritif au moins une fois par mois et non réfractaires au fromage. Au total, 240 femmes (âge moyen = 44 ans \pm 12) ont été interrogées selon le protocole de test hédonique traditionnel et 236 femmes (âge moyen = 42,7 ans \pm 11,8) ont été interrogées selon le protocole de choix expérimentaux non hypothétiques.

LES PRODUITS

Compte tenu de l'objectif de l'étude, l'efficacité de notre protocole de choix non-hypothétiques a été testée pour révéler des préférences relatives à une catégorie de produits peu impliquante. Les produits comparés dans cette étude sont deux marques de biscuits

feuilletés à l'emmental : une marque nationale (recette A) et une marque de distributeur (recette B). Ils sont présentés sans aucune information extrinsèque et codés respectivement X et Y.

PROCEDURE D'EVALUATION DU TEST HEDONIQUE TRADITIONNEL

Après s'être rincée la bouche avec de l'eau, chaque participante déguste et évalue les deux produits l'un à la suite de l'autre sur une échelle allant de 1 (« j'ai une très mauvaise opinion de ce biscuit ») à 10 (« j'ai une très bonne opinion de ce biscuit »)⁸. L'ordre de présentation des deux échantillons est équilibré sur l'ensemble du panel. Les sujets peuvent consommer autant de biscuits qu'ils le souhaitent mais doivent en consommer au moins un entier avant de déclarer leur appréciation globale. Après avoir noté les deux échantillons, la participante indique le biscuit qu'elle préfère. Une réponse « sans préférence » est possible.

PROCEDURE D'EVALUATION DU TEST DE CHOIX EXPERIMENTAUX NON-HYPOTHETIQUES

La procédure expérimentale du test de choix expérimentaux est scindée en deux étapes de mesures distinctes : la mesure de la préférence et la mesure du consentement à échanger.

Mesure de la préférence

La séance de test débute avec la dégustation des deux échantillons de biscuit en monadique séquentiel c'est-à-dire que les deux produits sont dégustés l'un à la suite de l'autre. La participante est informée dès le début du test qu'elle va avoir l'opportunité de choisir un lot d'un de ces deux biscuits à emporter chez elle. Une première proposition lui est donc présentée sur un bulletin (Figure 1) dans lequel on lui demande de choisir entre 300 g du biscuit X ou 300 g du biscuit Y (ce qui correspond à 3 paquets vendus dans le commerce). Ce premier choix permet de déterminer le biscuit préféré de la participante.

⁸ L'ensemble des études de ce travail de thèse ont été menées en utilisant cette formulation pour définir les bornes de l'échelle de mesure de l'appréciation globale.

Proposition1

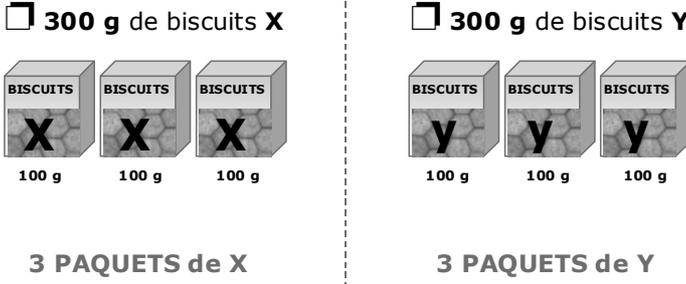
Pour commencer nous vous proposons de choisir entre deux lots de même quantité de biscuits de marque X ou de marque Y.

QUEL EST LE LOT QUI VOUS INTÉRESSE LE PLUS ?

300 g de biscuits X

300 g de biscuits Y

OU



3 PAQUETS de X

3 PAQUETS de Y

Figure 1 : Bulletin présentant la première alternative de choix

Mesure du consentement à échanger

La deuxième phase de la séance consiste à mesurer l'intensité de la préférence de la participante par le biais de son consentement à échanger son biscuit préféré pour une quantité supérieure du biscuit moins aimé. Pour cela, une série de bulletins proposant différentes alternatives de choix basées sur des variations de quantité des deux biscuits lui est présentée.

Les bulletins proposent systématiquement le biscuit préféré en quantité inférieure au biscuit moins aimé. Cependant, le consentement à échanger peut être évalué soit en diminuant la quantité du biscuit préféré, soit en augmentant la quantité du biscuit moins aimé. Il nous a paru intéressant d'utiliser ces deux types de mesures au niveau individuel pour vérifier si elles permettent d'aboutir aux mêmes estimations. Ainsi, deux séquences distinctes de choix sont mises en place dans un ordre équilibré sur l'ensemble de l'échantillon (la moitié des participantes débute par une séquence et l'autre moitié par l'autre séquence).

La séquence d'additions : la participante est face à des choix entre 300g du biscuit préféré contre des quantités croissantes du biscuit moins aimé allant de 350 g à 600 g avec un pas de 50 g. Cette séquence permet de déterminer la quantité minimale de produit supplémentaire nécessaire pour induire l'échange du biscuit préféré pour le biscuit moins aimé. Cette quantité minimale correspond à la valeur du consentement à échanger de la participante face à des augmentations de la quantité du biscuit moins aimé.

La séquence de réductions : la participante est face à des choix entre 300 g du biscuit moins aimé contre des quantités décroissantes du biscuit préféré allant de 250 g à 50 g avec un pas de 50 g. Cette séquence permet de déterminer la quantité minimale de produit à enlever pour induire l'échange du biscuit préféré pour le biscuit moins aimé. Cette quantité minimale

correspond à la valeur de consentement à échanger de la participante face à des réductions de la quantité du biscuit préféré.

Dans le cadre des deux séquences, au moment où la participante accepte d'échanger son biscuit préféré pour le biscuit moins aimé, le bulletin suivant permet d'accéder de manière plus précise aux valeurs de consentement à échanger avec un pas de 25g.

La figure 2 présente un exemple de bulletins et de réponses pour un participant préférant le produit X. Dans l'exemple des réponses présentées dans cette figure, les valeurs de consentement à échanger observées seraient les suivantes :

- séquence d'additions : consentement à échanger = + 200g
- séquence de réductions : consentement à échanger = - 150g.

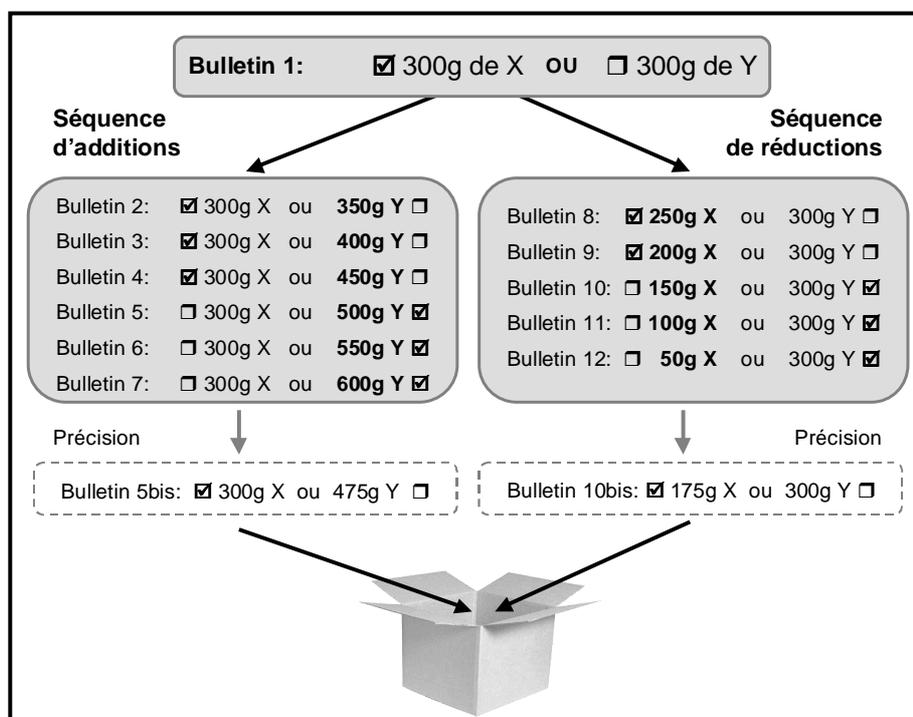


Figure 2 : Contenu des différentes alternatives de choix proposées à une participante préférant le biscuit X (bulletin 1) et débutant par la séquence d'additions (chaque ligne correspond à des choix proposés sur des bulletins individuels). Tous les bulletins sont placés dans une urne pour tirage au sort.

Les bulletins remplis par la participante sont déposés dans une urne au fur et à mesure de la séance. Dès le début du test, la participante est informée qu'un de ces bulletins sera tiré au sort à la fin de la séance et que le choix fait sur ce bulletin déterminera le produit et la quantité que la participante emportera chez elle. Ce tirage aléatoire permet de rendre l'ensemble des choix faits par la participante potentiellement effectifs. Le mécanisme incitatif de ce protocole est basée sur le fait que la participante ne doit pas regretter le choix qu'elle a

fait pour le bulletin tiré au sort et qu'elle a donc tout intérêt à répondre pour chaque bulletin en imaginant que ce soit celui qui sera tiré au sort.

A la fin de la séance, la participante répond à un ensemble de questions socio-démographiques dont les réponses seront utilisées pour interpréter son comportement d'échange.

RESULTATS ET DISCUSSION

TEST HEDONIQUE TRADITIONNEL

Préférence

Le pourcentage de participantes ayant déclaré préférer la recette A (54,4%) est comparé au pourcentage de participantes ayant déclaré préférer la recette B (45,6%) à l'aide d'un test binomial ($\alpha = 5\%$). Le résultat non significatif de ce test ($p = 0,103$) permet de conclure que la recette A n'est pas plus souvent préférée par les participantes que la recette B. Par ailleurs, quatorze participantes ont déclaré ne pas avoir de préférence et ne sont donc pas prises en compte dans l'analyse.

Intensité des préférences

L'intensité des préférences est calculée pour chaque individu par la différence entre la note hédonique donnée au produit préféré et la note hédonique donnée au produit moins aimé.

L'écart de note moyen des participantes préférant la recette A [moyenne $_{(A-B)} = 1,87$; écart-type = 1,97] est comparé à l'écart de note moyen des participantes préférant la recette B [moyenne $_{(B-A)} = 1,89$; écart-type = 1,73] à l'aide d'un test de Student non apparié ($t = 0,09$; $p = 0,926$). Ainsi, les participantes qui préfèrent la recette A ne présentent pas une différence d'appréciation supérieure aux participantes qui préfèrent la recette B.

TEST DE CHOIX EXPERIMENTAUX NON-HYPOTHETIQUES***Préférence***

L'analyse du premier choix montre que 60,2% des participantes choisissent d'emporter la recette A et que 39,8% choisissent d'emporter la recette B. Contrairement au résultat obtenu avec la question de préférence hypothétique du test hédonique traditionnel, l'écart est cette fois-ci très significatif (test binomial : $p < 0,001$) et montre qu'il y a plus de participantes qui préfèrent la recette A que de participantes qui préfèrent la recette B.

Intensité des préférences***Comparaison de l'intensité des préférences selon le biscuit préféré***

Les valeurs de consentement à échanger des participantes permettent de comparer l'intensité de la préférence des deux sous-populations révélées par le premier choix : les individus qui préfèrent la recette A (les pro-A) et les individus qui préfèrent la recette B (les pro-B). Cette comparaison est réalisée séparément sur les deux types de données collectées : la séquence d'additions et la séquence de réductions.

Séquence d'additions

La figure 3 illustre, pour les deux sous-populations, le pourcentage de participantes qui conservent leur biscuit préféré malgré les propositions croissantes de quantité supérieure du biscuit moins aimé.

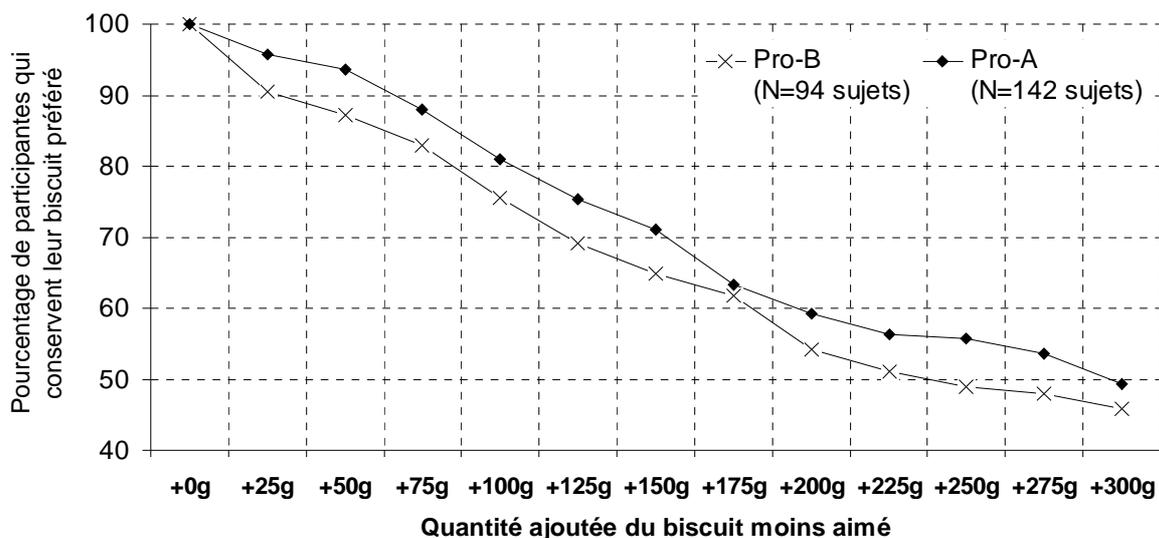


Figure 3 : Pourcentage de participantes qui conservent leur biscuit préféré en fonction des quantités croissantes proposées du biscuit moins aimé.

Dans un premier temps nous remarquons qu'un grand nombre de participantes conservent leur biscuit préféré quelles que soient les quantités du biscuit moins aimé ajoutées. Les participantes qui n'ont jamais échangé leur biscuit préféré contre le biscuit moins aimé seront considérées comme des participantes « fidèles » à leur biscuit préféré. Par opposition, les autres participantes ayant à un moment où un autre décidé de choisir le biscuit moins aimé seront appelées les participantes « opportunistes ». Il convient de comparer le pourcentage de participantes fidèles dans les deux sous-populations. Ainsi, 49% des pro-A sont fidèles à la recette A quelle que soit la quantité proposée de B et 46% des pro-B sont fidèles à la recette B quelle que soit la quantité de biscuits A proposée. Un test du χ^2 réalisé sur la répartition des individus fidèles ou opportunistes selon le produit préféré permet de conclure que cette répartition est indépendante du produit préféré ($\chi^2 = 0,29$; $p = 0,592$). Ce premier résultat suggère donc que l'intensité de la préférence des participantes est équivalente pour les deux biscuits.

Cependant, l'analyse plus approfondie des valeurs de consentement à échanger des participantes opportunistes qui échangent leur biscuit préféré à un moment donné met en évidence une supériorité de l'intensité de la préférence pour la recette A. Nous observons que la courbe correspondant aux pro-A est systématiquement au-dessus de la courbe représentant les pro-B indiquant donc que la préférence pour la recette A est légèrement plus stable que la préférence pour la recette B. Nous remarquons cependant que les deux sous-populations se comportent globalement de manière similaire avec une décroissance linéaire du pourcentage de participantes fidèles au fur et à mesure des propositions de plus en plus séduisantes. La principale différence entre ces deux courbes provient du comportement des individus face à la première proposition de 25 g de biscuits en plus. En effet, alors que 9,6 % des pro-B acceptent la proposition et échangent leur biscuit préféré, seulement 4,2 % des pro-A se laissent tenter par cette première proposition. Le comportement de ces quatorze participantes qui échangent leur biscuit préféré dès que le biscuit en compétition est proposé en quantité légèrement supérieure traduit donc une indifférence envers le type de biscuits à emporter et peut être assimilé au comportement des participantes qui déclarent ne pas avoir de préférence lors du test hédonique traditionnel.

Séquence de réductions

Les répartitions des valeurs de consentement à échanger sont analysées selon les deux sous-populations comme précédemment. La figure 4 représente le pourcentage de participantes qui conservent leur biscuit préféré en fonction des quantités déduites du biscuit préféré.

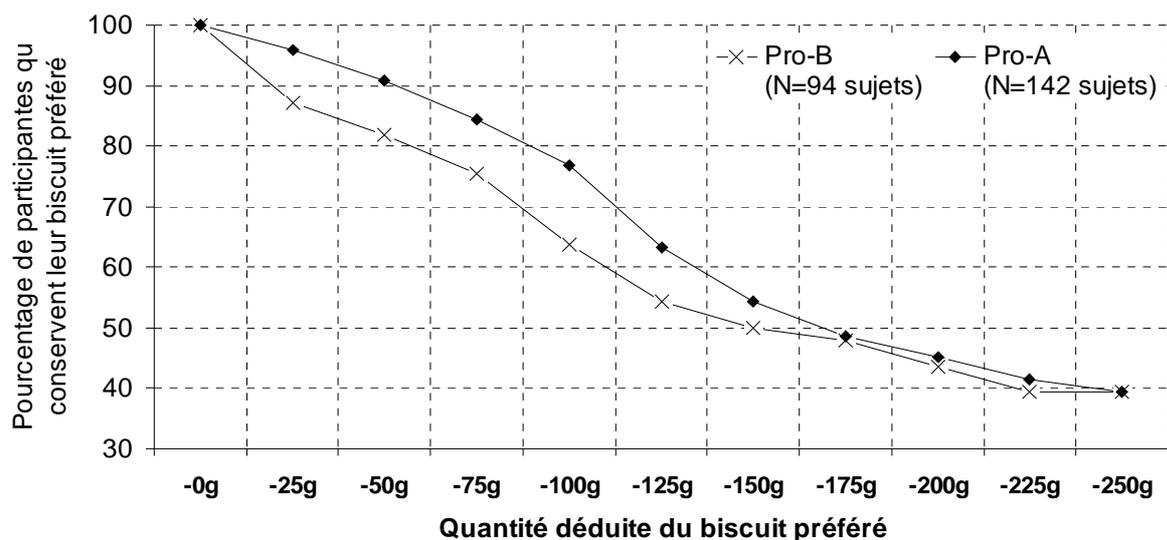


Figure 4 : Pourcentage de participantes qui conservent leur biscuit préféré en fonction des quantités déduites du biscuit préféré.

Comme le montre la figure 4, les deux sous-populations présentent strictement le même pourcentage de participantes fidèles (39%) ce qui illustre encore une fois que ce comportement est indépendant du produit préféré ($\chi^2 = 0,00$; $p = 0,991$). Cependant, l'analyse approfondie des valeurs de consentement à échanger démontre une différence de sensibilité des participantes opportunistes selon la sous-population. De la même manière que pour la séquence d'additions, la principale différence de comportement entre les deux sous-populations concerne la première réduction de 25 g du biscuit préféré. En effet, nous observons pour cette proposition que les pro-B sont plus facilement tentés par les lots de biscuits A que les pro-A par les lots de biscuits B. Cependant, contrairement à la séquence d'additions dans laquelle les deux sous-populations présentent ensuite un comportement similaire (courbes parallèles), nous observons qu'au-delà d'une réduction de 150 g, le pourcentage de pro-A qui conservent leur biscuit A devient équivalent au pourcentage de pro-B qui conservent leur biscuit B. Ainsi, même si les pro-A sont globalement plus fidèles que les pro-B, elles finissent quand même par accepter l'échange quand la quantité déduite est trop importante.

Simulation de la courbe de demande du biscuit A en compétition avec le biscuit B

Les données ont été présentées précédemment en analysant les comportements des pro-A et des pro-B séparément face aux différentes alternatives de choix. Cependant, chacun des choix fait par les deux sous-populations peut être ramené à l'ensemble de la population interrogée. Cette approche basée sur les comportements de l'ensemble des participantes est possible et permet d'estimer empiriquement l'évolution des parts de marché des deux produits concurrents dans un contexte de marché compétitif. Ainsi, les données obtenues pour les deux séquences de choix peuvent par exemple être analysées du point de vue du producteur du produit A. La figure 5 représente les conséquences sur les parts de marché de différentes offres du biscuit A face à une offre constante de 300 g de biscuits B au même prix.

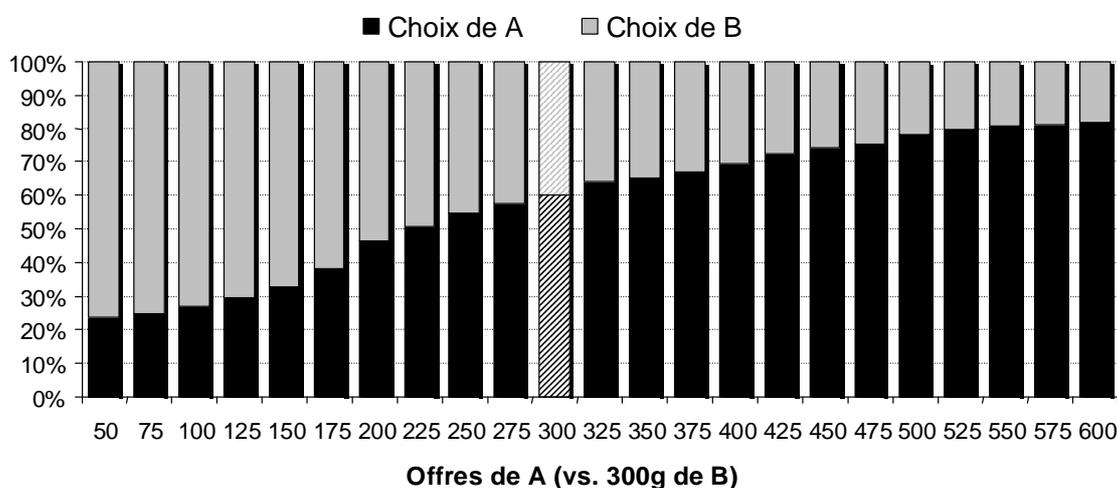


Figure 5 : Estimation empirique de l'évolution des parts de marché des deux produits.

Si les deux concurrents offrent la même quantité de biscuit (300 g) au même prix, alors 60,2% de la population choisit le biscuit A. Mais que se passe-t-il si la marque A décide de réduire la quantité de biscuit qu'elle propose dans ses paquets par exemple ? La partie gauche du graphique met en évidence les conséquences d'un tel scénario sur les parts de marché des deux biscuits. Cette partie du graphique est calculée sur la base des données de la sous-population des pro-A soumise à la séquence de réductions. Ainsi, pour chaque offre de biscuits A décroissante, il suffit d'éliminer de l'effectif précédent des pro-A les participantes qui décident de choisir le lot de biscuits B. Ainsi, nous remarquons que la quantité proposée de biscuits A peut être diminuée jusqu'à 250 g tout en restant choisie par plus de 50% de la population. A l'opposé, la partie droite du graphique illustre les conséquences positives qu'aurait une augmentation de la quantité de biscuits A sur les choix. Ici, ce sont donc les données de la sous-population des pro-B lors de la séquence d'additions qui sont utilisées. Il s'agit cette fois-ci pour chaque offre de A croissante d'éliminer de l'effectif précédent des pro-B les participantes qui décident de choisir le biscuit A.

Ce graphique basé sur des variations de quantité peut être utilisé de manière à évaluer la sensibilité du choix des individus par rapport à des variations de prix et ainsi aboutir à une courbe de demande. L'idée développée par Sanogo et Masters (2002) est d'utiliser les choix des participantes pour révéler leur consentement à payer pour les produits testés sur la base de la quantité nécessaire pour induire un échange. La valeur de WTP pour chaque produit est calculée sur la base de l'indifférence révélée par le consentement à échanger des participants. En effet, lorsque la participante accepte d'échanger le biscuit préféré pour le biscuit moins aimé, les deux lots présentés en quantités différentes lui apportent une satisfaction approximativement similaire. Ainsi, nous pouvons faire l'hypothèse que lors de l'alternative de choix qui entraîne un changement de biscuits :

$$p(A) \times q(A) \approx p(B) \times q(B)$$

avec p = prix du biscuit et q = quantité de biscuits

Au point d'échange du biscuit préféré pour le biscuit moins aimé, le rapport des prix des deux biscuits est donc l'inverse du rapport des quantités en concurrence. Si nous reprenons la figure 5, ce rapport de prix pour chaque alternative de biscuits A vs. 300 g de biscuits B peut être calculé et ainsi permettre d'évaluer les conséquences de variations de ce rapport de prix sur les parts de marché du biscuit A. Ces variations des parts de marché peuvent être assimilées à des variations de la demande du biscuit A en fonction du prix et la figure 6 représente donc la courbe de demande du biscuit A en compétition avec le biscuit B.

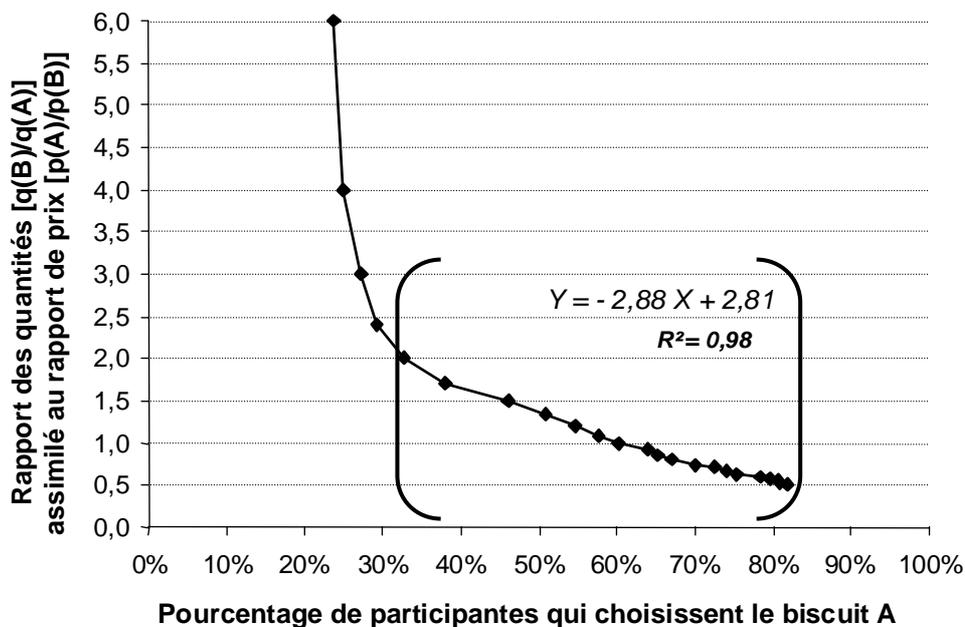


Figure 6 : Courbe de demande du biscuit A en fonction du prix relatif des deux produits et équation de la droite de régression calculée sur la partie linéaire.

L'observation de la partie gauche de cette courbe met en évidence les limites de notre hypothèse. Nous observons des parts de marché peu réalistes dans le cas des rapports de prix élevés entre les deux biscuits. Il est en effet fort peu probable que plus de 20% de la population continue d'acheter le biscuit A si celui-ci est vendu 6 fois plus cher que le biscuit B. Cependant, la partie linéaire de la courbe concernant les rapports de prix plus faibles (entre 0 et 2) peut être utilisée comme courbe de demande. L'équation de la droite de régression obtenue permet alors d'évaluer les conséquences de faibles variations de prix sur les parts de marché des deux biscuits A et B en compétition. Ainsi, si nous prenons par exemple le rapport de prix de ces deux marques de biscuits sur le marché ($p_{(A)}/p_{(B)} = 1,23$), l'équation de la droite de régression permet de prédire que, malgré ce rapport de prix, le biscuit A reste choisi par 55% de la population. Il serait intéressant de comparer les parts de marché obtenues par cette équation avec les parts de marché réelles des deux produits en compétition afin de révéler le rôle de l'appréciation de la recette seule par rapport aux variables marketing (publicité, marque) qui influencent l'achat de ces produits.

Comparaison des données obtenues selon les deux séquences de choix

Les valeurs de consentement à échanger obtenues selon la séquence d'additions ou selon la séquence de réductions peuvent être comparées sur l'ensemble de la population. La figure 7 représente le nombre de participantes qui conservent leur lot du biscuit préféré pour chaque offre du biscuit moins aimé en plus ou pour chaque réduction du biscuit préféré en moins.

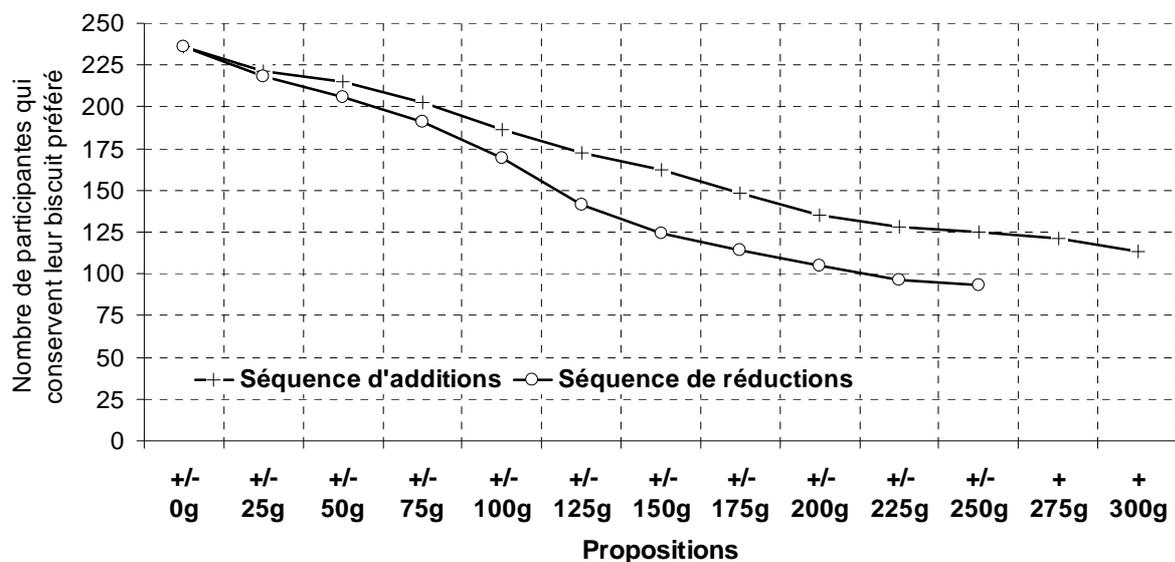


Figure 7 : Nombre de participantes qui conservent leur biscuit préféré en fonction des propositions (selon la séquence d'additions et selon la séquence de réductions).

Ce graphique met en évidence que plus la quantité du biscuit préféré/moins aimé est réduite/augmentée, plus le nombre de participantes qui décident d'accepter le lot du biscuit moins aimé à la place du lot du biscuit préféré est élevé. Cependant, nous remarquons qu'un grand nombre de participantes fidèles conservent le lot du biscuit préféré quelles que soient les propositions. Dans le cas de la séquence d'additions, 48% des participantes préfèrent repartir avec le lot de 300 g du biscuit préféré plutôt qu'accepter le lot de 600 g du biscuit moins aimé. Le nombre de participantes fidèles qui préfèrent, quant à elles, repartir avec 50 g du biscuit préféré au lieu de 300 g du biscuit moins aimé est légèrement moins important mais reste très significatif (40%).

La figure 7 montre également que la séquence d'additions aboutit à des valeurs de consentement à échanger globalement plus élevées que la séquence de réductions. Ainsi, la quantité minimum du biscuit le moins aimé à ajouter pour que 50% des participantes acceptent d'échanger leur biscuit préféré pour le biscuit le moins aimé est de 275 g, alors que la quantité minimum du biscuit préféré à enlever pour que 50% des participantes acceptent d'abandonner leur biscuit préféré est de 150 g. En d'autres termes, nous pouvons dire que les préférences sont moins stables face à des déductions du produit préféré que face à des augmentations du produit moins aimé. Ce phénomène peut sembler tout à fait cohérent étant donné que les enjeux lors des deux séquences ne sont pas les mêmes. Dans le cas des additions, les participantes sont certaines de repartir au pire avec 300 g de leur produit préféré alors que dans la phase de réduction, les participantes repartent au mieux avec 275 g de leur produit préféré. Par ailleurs, dans le cadre de la mesure des valeurs économiques privées, ce phénomène de divergence entre les valeurs de WTP et les valeurs de WTA pour le même bien est fréquent (Camerer, 1995 ; Horowitz et McConnell, 2002). Ce phénomène comportemental traduit en particulier le fait que les individus allouent plus de valeur à un bien s'ils le possèdent que s'ils ne le possèdent pas (Tversky et Kahneman, 1991). On peut le rapprocher de la « tendance à l'aversion pour les pertes » qui traduit une sensibilité aux pertes généralement plus élevée que la sensibilité à des gains d'un montant équivalent. Cette tendance, ajoutée au fait que la phase de déductions consiste en une dégradation d'une situation alors que la phase d'additions consiste en une amélioration d'une situation expliquent que les participantes ne déclarent pas des valeurs de consentement à échanger identiques selon les deux séquences. Malgré cette asymétrie, l'analyse de la relation entre les deux valeurs au niveau individuel démontre une nette cohérence de la part des participantes entre ces deux types de données. En effet, le coefficient de corrélation de Pearson (r) calculé entre les valeurs de consentement à échanger révélées lors des deux séquences pour les

participantes opportunistes est très significatif ($r = 0.717$; $p < 0.001$). Ainsi, un individu qui révèle un consentement à échanger élevé lors de la séquence d'additions révèle également un consentement à échanger élevé lors de la séquence de réductions.

Comportement d'échange : interprétation à l'aide des variables individuelles

Qui échange le biscuit préféré pour le biscuit moins aimé ?

Deux types de consommatrices ont été identifiés : les consommatrices fidèles qui n'échangent jamais leur biscuit préféré pour le biscuit moins aimé et les consommatrices opportunistes qui à un moment où à un autre finissent par accepter le lot du biscuit le moins aimé. Il nous a donc semblé intéressant de vérifier si certaines variables individuelles pourraient expliquer ce comportement de fidélité ou d'opportunisme. Pour cela, une régression logistique a été effectuée afin de déterminer la probabilité pour qu'une participante soit fidèle ou opportuniste en fonction de ses caractéristiques individuelles. Cette régression a été réalisée sur les comportements (fidèles ou opportunistes) observés pour les deux séquences qui sont dans ce cas considérées comme des répétitions. Les variables individuelles prises en compte dans le modèle sont les suivantes :

- la profession,
- l'âge : 18-34 ans ; 35-49 ans ; 50-65 ans,
- le niveau d'études : < baccalauréat ; baccalauréat ; > baccalauréat,
- la tranche de revenu du foyer : ≤ 2000 euros ; > 2000 euros,
- la taille du foyer : ≤ 2 ; > 2 ,
- le statut de consommatrice de la catégorie de produit : consommatrice ou non de biscuits feuilletés à l'emmental,
- le biscuit préféré : A ou B (réponse au choix 1),
- l'ordre de présentation des deux séquences : additions puis réductions ou réductions puis additions.

Cette régression logistique a mis en évidence qu'une seule des variables individuelles mesurées intervient de manière significative dans l'explication du comportement d'échange ou de non-échange des consommatrices. Il s'agit de la variable liée au statut d'utilisateur régulier de la catégorie de biscuit testée. Un test du χ^2 sur les données représentées dans le tableau 1 montre effectivement que le comportement fidèle ou opportuniste des

consommatrices est dépendant de leur statut de consommatrice de biscuits feuilletés à l'emmental ($\chi^2 = 5,49$; $p = 0,064$).

Tableau 1 : Répartition des participantes selon leur statut de consommatrice de la catégorie de biscuit

	Fidèle	Opportuniste sur une séquence	Opportuniste sur les deux séquences	Somme
<i>Non consommatrices</i>	51	25	86	162
<i>Consommatrices</i>	35	9	30	74

Les consommatrices qui déclarent spontanément consommer régulièrement ce type de biscuits tendent à être plus fidèles que les consommatrices qui ne déclarent pas consommer ce type de biscuits. Ainsi, les consommatrices régulières attachent plus d'importance à la qualité du lot qu'elles emporteront chez elles que les non-consommatrices. Lange *et al.* (2002) ont observé un comportement similaire de la part de consommateurs fidèles à une marque de Champagne qui révèlent des valeurs de WTP non hypothétiques plus élevés que des consommateurs n'ayant pas de fidélité envers une marque. Ainsi, le comportement de fidélité observé dans cette étude ne reflèterait pas uniquement la qualité perçue par les individus mais également la familiarité des participants avec la catégorie de produit testée.

Qu'est ce qui influence le niveau des valeurs de consentement à échanger ?

Deux analyses de la variance (ANOVA) à 8 facteurs (*consentement à échanger = biscuit préféré + ordre de présentation des séquences + profession + âge + niveau d'études + revenu du foyer + taille du foyer + statut de consommatrice de la catégorie*) sont réalisées sur les données de consentement à échanger des participantes opportunistes de chaque séquence afin de déterminer si certaines variables individuelles peuvent expliquer un niveau de consentement à échanger plus ou moins élevé.

Séquence d'additions

Le tableau 2 présente les résultats de l'ANOVA sur les valeurs de consentement à échanger de la séquence d'additions. Même si la figure 3 présentée précédemment montrait une légère supériorité des valeurs de consentement à échanger des pro-A sur celles des pro-B, nous observons ici que les moyennes des deux sous-populations ne sont pas significativement différentes.

Tableau 2 : Résultats de l'ANOVA sur les consentements à échanger de la séquence d'additions (les variables en gras sont significatives au seuil de 10 %).

Source	ddl	F de Fisher	Pr > F
Biscuit préféré	1	1,62	0,206
Ordre des séquences	1	3,09	0,081
Profession	3	0,85	0,468
Niveau d'études	2	4,93	0,009
Statut de consommatrice	1	0,01	0,940
Revenu du foyer	2	1,92	0,152
Age	2	4,22	0,017
Taille du foyer	1	0,04	0,840

La figure 8 illustre l'impact des variables individuelles significatives sur les valeurs de consentement à échanger par rapport à la moyenne du groupe.

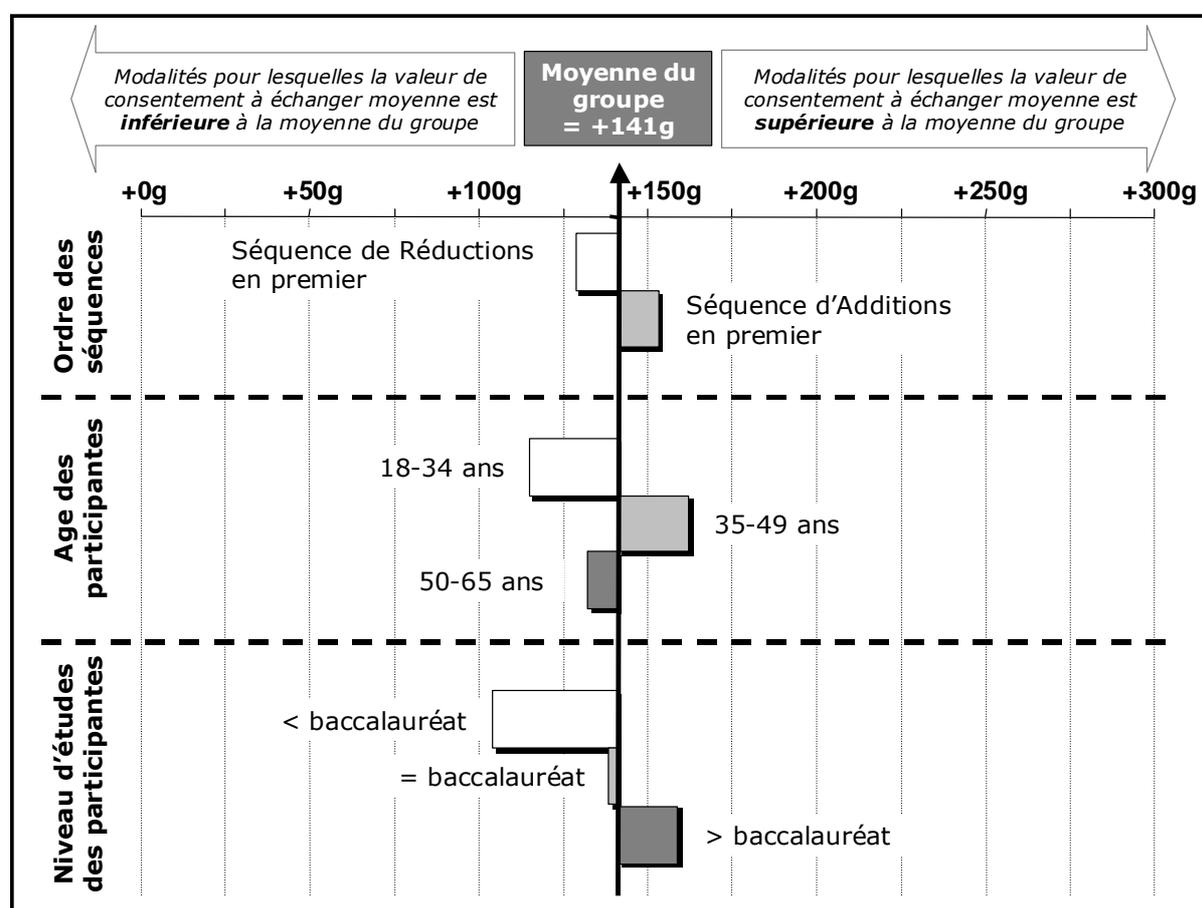


Figure 8 : Illustration de l'impact des variables individuelles significatives (ordre des séquences, âge et niveau d'études des participantes) sur le niveau des valeurs de consentement à échanger face à des additions du biscuit moins aimé.

L'ordre de présentation des deux séquences a tendance à influencer le niveau des valeurs de consentement à échanger. Les participantes ayant débuté la séance par la séquence d'additions révèlent des valeurs de consentement à échanger face à des additions globalement

supérieures aux participantes ayant débuté la séance par la séquence de réductions. Ceci peut s'expliquer par le fait que lors de la phase de réductions, les participantes échangent leur biscuit plus facilement que lors de la phase d'additions. Ainsi, si la séquence d'additions est précédée de la séquence de réductions, les participantes souhaitent rester cohérentes et échangent donc plus facilement leur produit que si elles n'ont pas été soumises à la séquence de réductions précédemment.

En ce qui concerne les variables liées au statut de l'individu et non au protocole, seuls le niveau d'études et l'âge des participantes influencent significativement le niveau des valeurs de consentement à échanger. Nous observons sur la figure 8 que les personnes les plus jeunes et les plus âgées présentent des valeurs de consentements à échanger plus faibles que les participantes entre 35 et 49 ans. Enfin, l'impact du niveau d'études met en évidence que plus les participantes ont un niveau d'études élevé, plus elles ont tendance à présenter des valeurs de consentement à échanger élevées. Même si le revenu du foyer n'influence pas significativement les valeurs de consentement à échanger, nous pouvons supposer que le niveau d'études traduit indirectement le niveau social de l'individu et donc que les personnes d'un niveau social plus faible sont plus sensibles aux promotions.

L'influence significative de certaines variables liées au statut des individus (âge ou niveau d'étude) ou au protocole (ordre de présentation des séquences) suggère que le comportement d'échange face à des additions du produit moins aimé peut ne pas refléter uniquement une intensité de préférence envers un produit mais également un comportement général des individus face à des promotions (ou un rapport à l'argent).

Séquence de réductions

Le tableau 3 présente les résultats de L'ANOVA réalisée sur les données de la séquence de réductions.

Tableau 3 : Résultats de l'ANOVA sur les consentements à échanger de la séquence de réductions (les variables en gras sont significatives au seuil de 10%)

Source	ddl	F de Fisher	Pr > F
Biscuit préféré	1	5,48	0,021
Ordre des séquences	1	1,11	0,294
Profession	3	0,09	0,966
Niveau d'études	2	5,41	0,006
Statut de consommatrice	1	0,17	0,682
Revenu du foyer	2	0,03	0,967
Age	2	0,83	0,436
Taille du foyer	1	1,29	0,258

Les résultats sont différents de ceux de l'ANOVA réalisée sur les données de la séquence d'additions. L'effet « *biscuit préféré* » est significatif ce que l'interprétation de la figure 4 laissait supposer : l'intensité de la préférence des pro-A est supérieure à celle des pro-B. En ce qui concerne les variables liées au statut de l'individu, nous retrouvons l'effet significatif du niveau d'étude avec une augmentation des valeurs de consentement à échanger en fonction de l'augmentation du niveau d'études. En revanche, l'ordre de présentation des séquences ainsi que l'âge des interviewés n'interviennent plus de manière significative sur les valeurs de consentement à échanger face à des réductions.

La figure 9 illustre l'impact des deux variables significatives sur le niveau de consentement à échanger des participantes par rapport à la moyenne du groupe.

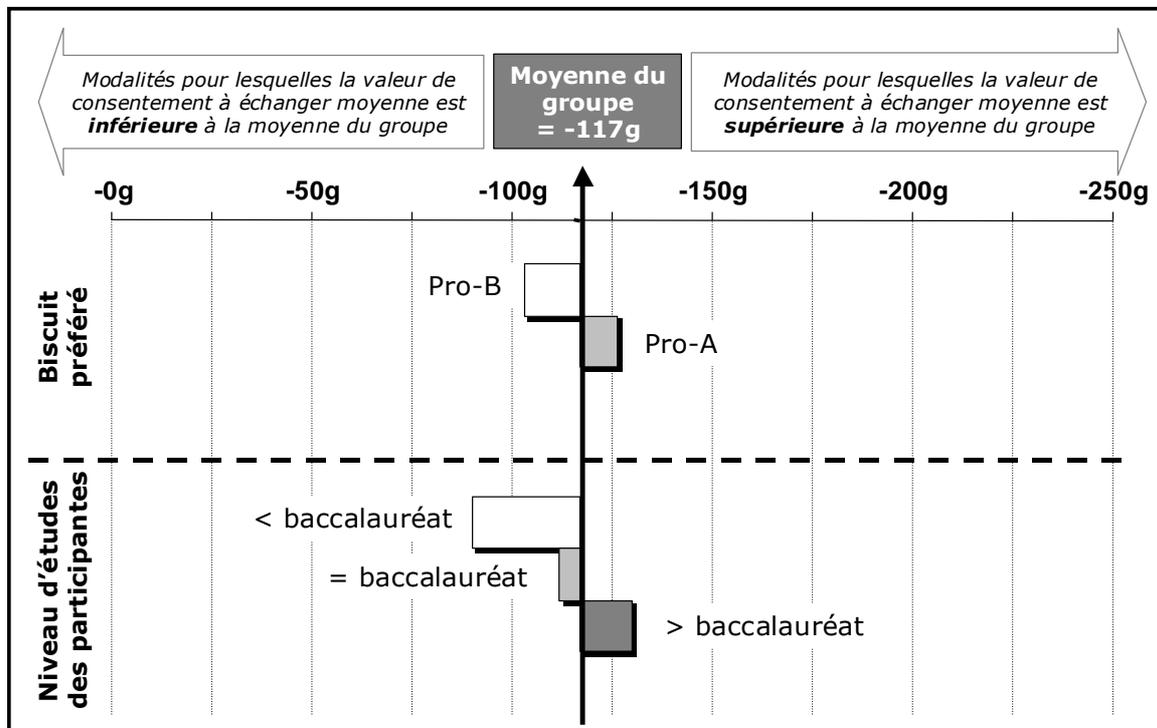


Figure 9 : Illustration de l'impact des variables individuelles significatives (*biscuit préféré* et *niveau d'études*) sur le niveau des valeurs de consentement à échanger face à des réductions du biscuit préféré.

Dans une perspective de comparaison des deux séquences de choix, nous pouvons faire l'hypothèse que la séquence de réductions est plus efficace que la séquence d'additions pour recueillir des données de préférence focalisées sur les produits dégustés. En effet, les valeurs de consentement à échanger observées face à des réductions sont peut être plus représentatives d'une intensité de préférence envers les produits testés que ne le sont les valeurs de consentement à échanger observées face à des additions qui illustreraient peut-être plus un comportement général des individus face à un compromis qualité/quantité.

DISCUSSION GENERALE

La mise en place d'un tel protocole de choix expérimentaux non-hypothétiques dans le cadre habituel des études consommateurs quantitatives n'a, à notre connaissance, jamais été réalisé. En effet, ces méthodologies sont principalement développées dans le cadre de recherches académiques avec des protocoles relativement lourds. L'objectif de ce travail était de vérifier les possibilités techniques et budgétaires de mise en œuvre d'une telle méthodologie par des enquêteurs non expérimentés lors d'un test hédonique en salle avec des participants recrutés ponctuellement dans la rue. Outre l'analyse de faisabilité, la présente étude a également permis de comparer les résultats obtenus à l'aide de ce protocole à ceux d'un test hédonique traditionnel. Ce travail a donc permis de mettre en évidence les forces et les faiblesses de l'approche méthodologique non hypothétique par rapport à un protocole de test traditionnel.

APPORTS DE CETTE APPROCHE METHODOLOGIQUE

Importance d'une procédure non-hypothétique

La comparaison du résultat de la question de préférence posée lors du test hédonique traditionnel avec celui obtenu suite à la première proposition du test de choix expérimentaux démontre l'importance de l'implication des participants dans leurs réponses. En effet, la préférence hypothétique du test traditionnel ne permet pas de discriminer les deux biscuits alors que la question portant sur le choix réel de biscuits à emporter révèle une nette préférence pour le biscuit A. Ainsi, l'idée de motiver les interviewés en leur offrant le produit qu'ils déclarent préférer est déjà une amélioration qui serait assez simple à mettre en œuvre pour révéler les vraies préférences dans le cadre d'études consommateurs sur des aliments de consommation courante de faible valeur monétaire. Cependant cette approche reste difficile à réaliser dans le cadre de l'évaluation des préférences envers des aliments plus onéreux comme du champagne ou du vin pour lesquels les méthodes d'enchères plus classiques basées sur l'achat des produits avec de l'argent sont probablement mieux adaptées et moins coûteuses.

Richesse de l'information supplémentaire apportée par les séquences de choix basées sur des variations des quantités offertes

Même si le simple choix entre deux lots de produit à emporter (choix 1) pourrait être appliqué sans avoir recours aux mesures de consentement à échanger pour révéler la préférence du sujet, les choix successifs basés sur les variations des quantités offertes se justifient pleinement pour plusieurs raisons.

Premièrement, ils permettent de confirmer la préférence révélée par le premier choix et de corriger en partie les biais liés à une simple question de choix d'un produit à emporter. En effet, ce premier choix reste critiquable du fait qu'il peut ne pas être réalisé en accord avec la préférence de l'interviewé. Un produit peut être choisi par l'interviewé pour faire plaisir à un autre membre de sa famille. La curiosité d'emporter le produit le plus nouveau, le plus étrange peut également faire basculer le participant vers un choix différent de sa préférence.

Par ailleurs, les choix successifs permettent d'obtenir des informations comportementales individuelles particulièrement riches. Nous avons ainsi pu observer que dans la population interviewée, certains individus restent fidèles à leur produit quel que soit leur premier choix alors que d'autres au contraire n'ont pas de réelle préférence et échangent leur (soit disant) produit préféré dès que l'autre produit est présenté en quantité légèrement supérieure.

Les différentes valeurs de consentement à échanger calculées sur la base des choix successifs ont également permis de comparer l'intensité des préférences des individus et nous avons ainsi pu mettre en évidence une meilleure discrimination des deux sous-populations des pro-A et des pro-B par l'intermédiaire des valeurs de consentement à échanger que par l'intermédiaire des écarts de notes hédoniques du test traditionnel. De la même façon, Lange *et al.* (2002) ont également observé pour du Champagne que des valeurs de WTP non hypothétiques aboutissent à une meilleure discrimination des différents produits que des scores traditionnels d'appréciation hédonique.

Enfin, nous avons également montré l'utilité opérationnelle de ce type de données concernant les réponses des participants aux différentes alternatives de choix de quantités variables. Ainsi, nous avons pu estimer les conséquences sur les parts de marché de différents scénarios de production comme la réduction de la quantité de biscuits proposée dans un lot. Nous pourrions de la même façon estimer les conséquences qu'aurait une action de promotion réalisée par l'entreprise concurrente. Enfin, ces données nous ont également permis d'estimer une courbe de demande du produit A en compétition avec le produit B pouvant être utilisée

pour évaluer les conséquences de variations de prix de vente sur les parts de marché des deux produits mis en concurrence.

Considérations méthodologiques

Un budget de test équivalent à celui d'un test hédonique traditionnel en salle a été respecté. La seule dépense supplémentaire concerne le coût des produits donnés aux participants, mais il reste tout à fait raisonnable dans le cadre de produits de faible valeur monétaire.

En ce qui concerne les enquêteurs, cette nouvelle approche méthodologique est apparue motivante et agréable à administrer. Par ailleurs, le recrutement des sujets a été largement facilité par l'annonce d'un lot de produits à emporter à la maison.

Les participantes ont également montré une disponibilité et une écoute supérieures à celles dont elles font preuve habituellement. La durée des séances de test est globalement équivalente à celle d'un test traditionnel et, contrairement aux séances de test plus classique, les participantes n'ont pas exprimé de sentiment négatif envers la longueur de la séance. L'ambiance était globalement beaucoup plus détendue (enquêteurs et interviewés) que lors des études plus traditionnelles.

LIMITES DE CETTE APPROCHE METHODOLOGIQUE

Considérations méthodologiques

Certaines difficultés ont cependant été rencontrées en ce qui concerne la mise en œuvre par des enquêteurs non familiers de ce type d'approche. En effet, le démarrage de l'étude a été compliqué et a nécessité un briefing relativement long des enquêteurs. Il a finalement fallu que tous les enquêteurs soient soumis au test afin d'être certain qu'ils comprennent les mécanismes mis en œuvre dans cette approche. Par la suite, les enquêteurs ont eu tendance à s'éloigner du discours explicatif et à employer des mots indésirables comme « jeu » ou « loterie » qui peuvent faire naître chez le participant des comportements de jeu ou de prise de risque non souhaités.

Des comportements inattendus

Certaines participantes ont manifesté des comportements inattendus. Quatre participantes ont réalisé des choix sans aucune logique apparente c'est-à-dire en choisissant l'un ou l'autre des biscuits quelles que soient les propositions de choix. Les données de ces

participantes ont été supprimées de l'analyse. Ces réponses versatiles peuvent provenir d'une difficulté de compréhension de la tâche pour certains interviewés et appellent donc à un travail sur la façon d'expliquer le test aux participants. Cependant, elles peuvent également refléter un manque d'implication du participant dans le test. Ce retour possible sur la façon dont a été vécu le test par les participants est alors particulièrement intéressant puisqu'il permet d'éliminer les participants non impliqués dans le test, ce qu'un test hédonique traditionnel ne permet pas.

D'autre part, dans le cas de la séquence d'additions, quinze participantes présentent un comportement d'échange que nous n'avions pas anticipé. Après avoir réalisé un échange de leurs 300 g de produit préféré pour une quantité supérieure du produit moins aimé, les offres toujours croissantes mettent alors en évidence un nouvel échange vers le produit préféré lorsque la quantité proposée du produit moins aimé est très importante (5 ou 6 paquets). Ce comportement semble être indépendant de la préférence envers l'un ou l'autre des biscuits et semble plutôt lié à une trop grande quantité de biscuit proposée. Ainsi, pour ces participantes, les valeurs de consentement à échanger retenues correspondent à leur premier échange vers le produit moins aimé. Ce comportement provient du fait que, contrairement à l'argent, l'utilité pour le participant peut décroître au-delà d'une certaine quantité de biscuits. En effet, des facteurs autres que l'appréciation du produit peuvent intervenir dans le choix entre les deux lots lorsque les quantités proposées sont importantes. Le participant peut ne pas souhaiter s'embarasser d'une quantité de produit volumineuse pendant la suite de sa journée ou même chez lui. La recherche de diversité peut également se manifester chez le participant en cherchant à éviter de devoir consommer de nombreux paquets du même biscuit avant la date de péremption. Ce phénomène de décroissance de l'utilité pour des quantités trop importantes doit donc être pris en compte dans la détermination de la quantité initiale de produit à proposer si l'on souhaite conserver l'attraction pour une quantité supérieure de produit moins aimé.

Enfin, même si ce processus de troc semble efficace pour évaluer l'attachement des participants à leur produit préféré dans le cas de produits à faible valeur monétaire comme des biscuits apéritifs, nous ne nous attendions pas à ce qu'autant de participantes restent fidèles à leur produit préféré quelle que soit l'alternative de choix. Nous avons interprété ce comportement de non-échange comme une fidélité totale envers le biscuit préféré et donc en une préférence très élevée. Cependant, le fait que le nombre de participantes qui ne se laissent jamais tenter par le lot du biscuit moins aimé soit strictement équivalent dans les deux sous-populations peut également suggérer que ce phénomène de non-échange corresponde à une

manifestation de désirabilité sociale. En effet, ces participantes peuvent simplement ne pas souhaiter paraître opportunistes face aux enquêteurs et montrer qu'elles privilégient la qualité et non la quantité. Enfin, certaines personnes fuient l'inconfort mental généré par la nécessité de prendre une décision. Une fois que ces personnes ont réussi à faire leur premier choix, elles s'y tiennent jusqu'à la fin du test. Il serait donc important de pouvoir distinguer les comportements de non-échange liés à une très forte préférence de ceux liés à un comportement de désirabilité sociale ou de refus de l'inconfort mental.

PERSPECTIVES D'AMELIORATIONS

Cette première étude étant exploratoire, nous avons fait le choix de mettre en œuvre deux mesures de consentements à échanger selon deux séquences de troc différentes. Dans un objectif d'application de cette méthode en routine, il n'est pas indispensable de réaliser ces deux mesures simultanément. Même si la séquence d'additions peut sembler plus naturelle pour les participants qui sont quotidiennement confrontés à des promotions, l'analyse des données de chacune des séquences de troc a démontré une supériorité de la séquence de réductions sur la séquence d'additions concernant plusieurs points. Premièrement, la fidélité des participantes est plus rudement mise à l'épreuve lors de la séquence de réductions dans laquelle les conséquences cognitives négatives sont plus importantes que dans le cas de la séquence d'additions. Ce phénomène se traduit donc par des comportements d'opportunisme plus nombreux et donc un nombre d'observations de consentement à échanger plus important dans la séquence de réductions que dans la séquence d'additions. D'autre part, les valeurs de consentement à échanger révélées par la séquence de réductions permettent de discriminer de manière significative les deux sous-populations alors que les données de la séquence d'additions ne le permettent pas. Enfin, la séquence de réductions empêche les problèmes liés à une trop grande quantité de produit proposée qui se manifestent dans la séquence d'additions.

Pour conclure, cette approche de troc basée sur des quantités de produits à échanger est facile à mettre en place lors d'études quantitatives sur des produits alimentaires de faible valeur monétaire. Cependant, même si cette méthodologie semble tout à fait pertinente dans un objectif de positionnement d'un produit en termes de performance organoleptique par rapport à son concurrent principal, il reste encore des voies de travail à explorer de manière à permettre la comparaison de plus de deux produits. De la même façon, la mise en place d'une

telle approche dans le cadre de produits frais peut également être problématique étant donné que le consommateur recruté ponctuellement ne rentre pas systématiquement chez lui après la séance de test. Un système de distribution des produits gagnés pourrait éventuellement être mis en place mais aurait un fort impact sur le coût de l'étude. Savoir comment adapter ce protocole de test lors de la comparaison de plus de deux produits ou de produits frais sont autant de questions qui n'ont pas été résolues dans ce travail et qui mériteraient donc d'être approfondies afin de pouvoir utiliser en routine cette mesure du consentement à échanger.

RÉFÉRENCES

- Becker G.M., DeGroot M.H. et Marschak J. (1964) - Measuring utility by a single-response sequential method, *Behavioral Science*, volume 9, p. 226-232.
- Camerer C. (1995) - Individual decision making, *The handbook of experimental economics*, J.H. Kagel et A.E. Roth, Princeton, Princeton University Press.
- Esposito J. et Jobe J. (1991) - A general model of the survey interaction process, *Seventh Annual Research Conference of the Bureau of the Census*, July, Washington, DC.
- Fox J.A., Shogren J.F., Hayes D.J. et Kliebenstein J.B. (1998) - CVM-X: calibrating contingent values with experimental auction markets, *American Journal of Agricultural Economics*, volume 80, p. 455-465.
- Hoffman E., Menkhous D.J., Chakravarti D., Field R.A. et Whipple G.D. (1993) - Using laboratory experimental auctions in marketing research: a case study of new packaging for fresh beef, *Marketing Science*, volume 12, p. 318-338.
- Horowitz J.K. et McConnell K.E. (2002) - A review of WTA/WTP studies, *Journal of Environmental Economics and Management*, volume 44, p. 426-447.
- Jaeger S.R., Lusk J.L., House L.O., Valli C., Moore M., Morrow B. et Traill W.B. (2004) - The use of non-hypothetical experimental markets for measuring the acceptance of genetically modified foods, *Food Quality and Preference*, volume 15, p. 701-714.
- Kalwani M.U. et Silk A.J. (1982) - On the reliability and predictive validity of purchase intention measures, *Marketing Science*, volume 1, p. 243-286.
- Krosnick J. (1991) - Response strategies for coping with the cognitive demands of attitude measures in surveys, *Applied Cognitive Psychology*, volume 5, p. 213-236.
- Lange C., Martin C., Chabanet C., Combris P. et Issanchou S. (2002) - Impact of the information provided to consumers on their willingness to pay for champagne: comparison with hedonic scores, *Food Quality and Preference*, volume 13, p. 597-608.
- Leon F., Couronne T., Marcuz M.C. et Koster E.P. (1999) - Measuring food liking in children: a comparison of non verbal methods, *Food Quality and Preference*, volume 10, p. 93-100.
- Levy C.M. et Koster E.P. (1999) - The relevance of initial hedonic judgements in the prediction of subtle food choices, *Food Quality and Preference*, volume 10, p. 185-200.
- List J.A. et Shogren J.F. (1998) - Calibration of the differences between actual and hypothetical valuations in a field experiment, *Journal of Economic Behavior and Organization*, volume 37, p. 193-205.
- Lusk J.L. (2003) - Effect of cheap talk on consumer willingness to pay for Golden rice, *American Journal of Agricultural Economics*, volume 85, p. 840-856.
- Lusk J.L. et Schroeder T.C. (2004) - Are choice experiments incentive compatible? A test with quality differentiated beef steaks, *American Journal of Agricultural Economics*, volume 86, p. 467-482.
- Marette S., Roosen J., Blanchemanche S. et Verger P. (2006) - Health information and the choice of fish species: an experiment measuring the impact of risk and benefit information, *Working Paper 06-WP 421* (unpublished: www.card.iastate.edu).
- Peryam D.R., Polemis B.W., Kamen J., Eindhoven J. et Pilgrim F.J. (1960) - *Food preferences of men in the armed forces*, Chicago, Quatermaster Food and Container Institute.

Sanogo D. et Masters, W.A. (2002) - A market-based approach to child nutrition: mothers' demand for quality certification of infant foods in Bamako, Mali, *Food Policy*, volume 27, p. 251-268.

Schutz H.G. (1957) - Preference ratings as predictors of food consumption, *American Psychologist*, volume 12, p. 380

Simon H.A. (1956) - Rational choice and the structure of the environment, *Psychological Review*, volume 63, p. 129-152.

Stefani G., Romano D. et Cavicchi A. (2006) - Consumer expectations, liking and willingness to pay for specialty foods: do sensory characteristics

tell the whole story? *Food Quality and Preference*, volume 17, p. 53-62.

Stolte J.F. (1994) - The context of satisficing in vignette research, *The Journal of Social Psychology*, volume 134, p. 727-733.

Tversky A. et Kahneman D. (1991) - Loss aversion in riskless choice: a reference-dependent model, *Quarterly Journal of Economics*, volume 106, p. 1039-1061.

Vickrey W. (1961) - Counterspeculation, auctions, and competitive sealed tenders, *Journal of Finance*, volume 16, p. 8-37.

Chapitre III

QUESTIONS SOULEVEES ET ELEMENTS DE REPONSES

Ce travail sur les méthodologies de test consommateurs a soulevé de nombreuses questions tant au niveau de leur mise en œuvre qu'au niveau de l'interprétation de l'information hédonique récoltée. Ce dernier chapitre a pour objectif de présenter des éléments de réponses aux différentes questions auxquelles nous avons été confrontés pendant ce travail de thèse. Nous verrons cependant que certains points abordés dans ce travail restent encore non résolus. Pour terminer, nous envisagerons l'utilisation des propositions méthodologiques développées dans un contexte industriel routinier.

A - QUESTIONS METHODOLOGIQUES

L'objectif de la première partie expérimentale de ce travail était d'illustrer les risques pris lors de l'utilisation d'un test en conditions artificielles de dégustation à la place d'un test en conditions naturelles de consommation. Cet objectif a nécessité la réalisation de plusieurs tests (en salle et à domicile) avec l'aide d'instituts d'études. La mise en place de ces expérimentations a soulevé plusieurs questions méthodologiques pour lesquelles je pense maintenant avoir plus de recul et que je souhaiterais aborder.

I - NOTE HEDONIQUE ET/OU PREFERENCE ?

Outre les notes hédoniques obtenues au fur et à mesure des dégustations, les tests en salle ou à domicile proposent généralement, à la fin du test, une question de préférence ou de classement. Ainsi, alors que la note hédonique semble aujourd'hui clairement la mesure sur laquelle les conclusions statistiques sont obtenues, les données de la question de préférence sont parfois utilisées pour communiquer les résultats de manière plus simple et plus percutante.

La comparaison de ces deux types de questions n'a pas été détaillée dans ce manuscrit mais généralement les résultats entre les notes hédoniques et la question de préférence lors des tests en salle que nous avons menés sont plutôt cohérents. Nous pouvons donc nous poser la question de l'utilité de ces deux types de mesures lors d'un test en salle si ce n'est, peut-être, pour tester la cohérence des sujets et éliminer les participants qui classent les produits différemment selon les deux questions (ce qui correspond à 5% des sujets dans le cas des études menées dans ce travail). En revanche, la cohérence individuelle note hédonique/préférence lors des tests à domicile est moins bonne (10-15 % de sujets incohérents). Ce phénomène n'est cependant pas très surprenant puisque la préférence entre

les deux produits testés fait intervenir une variable « mémoire » étant donné que le sujet doit se souvenir de ce qu'il a testé la semaine précédente. Récolter l'appréciation des recettes basée sur la mémorisation de l'expérience avec le produit serait pertinent puisque c'est finalement cette appréciation mémorisée et non l'appréciation instantanée qui intervient dans le processus d'achat. Cependant, la façon dont cette préférence est obtenue aujourd'hui dans les tests à domicile (c'est-à-dire juste après la note hédonique donnée au deuxième produit) n'est pas optimale étant donné que seule l'évaluation du premier produit dégusté est basée sur une expérience mémorisée. La comparaison entre les deux produits testés n'est donc pas complètement équitable et il serait intéressant de mettre en place un système de rappel téléphonique quelques semaines après le test pour demander aux sujets leur préférence concernant les produits auxquels ils ont été confrontés.

II - EFFECTIF DU PANEL

Les tests en salle et à domicile réalisés sur les laits fermentés, les eaux gazeuses et les biscuits d'apéritif ont été menés selon un design en monadique séquentiel avec un ordre de présentation équilibré sur l'ensemble du panel. L'analyse des données, à la fois en monadique séquentiel (sur les deux produits) et en monadique pur (données du premier produit dégusté) a nécessité de travailler avec des effectifs de panels supérieurs à ce qui est habituellement mis en œuvre chez Danone (c'est-à-dire 120 sujets avec une tolérance jusqu'à 80 sujets au minimum). Ainsi, afin d'avoir une lecture des données en monadique pur basée sur ces effectifs, les tests ont été menés avec 240 ou 160 sujets. Ces effectifs élevés nous ont permis d'appliquer notre outil LCO (Leave Consumers Out) présenté dans l'article 5 (annexe I) pour tester la robustesse des données des tests à domicile réalisés lors de ces trois expérimentations. Nous avons ainsi pu tester la robustesse de la différence significative d'appréciation obtenue dans ces trois études face à la diminution de la taille du panel. La figure 26 présente, pour chaque étude, l'évolution du pourcentage de panels aléatoires pour lesquels la différence d'appréciation entre les deux produits reste significative.

Ces graphiques illustrent clairement l'impact de la taille du panel interrogé sur les conclusions obtenues et soulèvent donc bien la nécessité de travailler sur un effectif de sujets important. Cependant, l'utilisation de ces graphiques pour aider à la détermination d'un nombre de sujets minimum à interroger reste compliquée. En effet, si nous nous basons sur l'effectif de 120 sujets souvent recommandé par les statisticiens, nous observons que les conséquences sur la conclusion obtenue sont différentes selon les études. Alors que l'étude sur les biscuits d'apéritif présente une différence d'appréciation stable à partir de 120 sujets,

la différence d'appréciation observée pour les laits fermentés et les eaux gazeuses aurait eu moins d'une chance sur deux d'être observée si nous avions interrogé 120 sujets. Dans le cas des études sur les laits fermentés et sur les eaux gazeuses, les courbes ne présentent même pas de plateau indiquant une stabilisation de la conclusion à partir d'un certain effectif ce qui laisse supposer que nous aurions pu tout à fait ne pas obtenir cette différence significative d'appréciation avec un autre groupe de sujets de la même taille. Quoi qu'il en soit, ces résultats condamnent clairement l'utilisation d'un effectif minimum de 60 sujets proposé par la norme XP V09-500 (AFNOR, 2000) qui, dans tous les cas, est insuffisant pour obtenir une différence d'appréciation robuste.

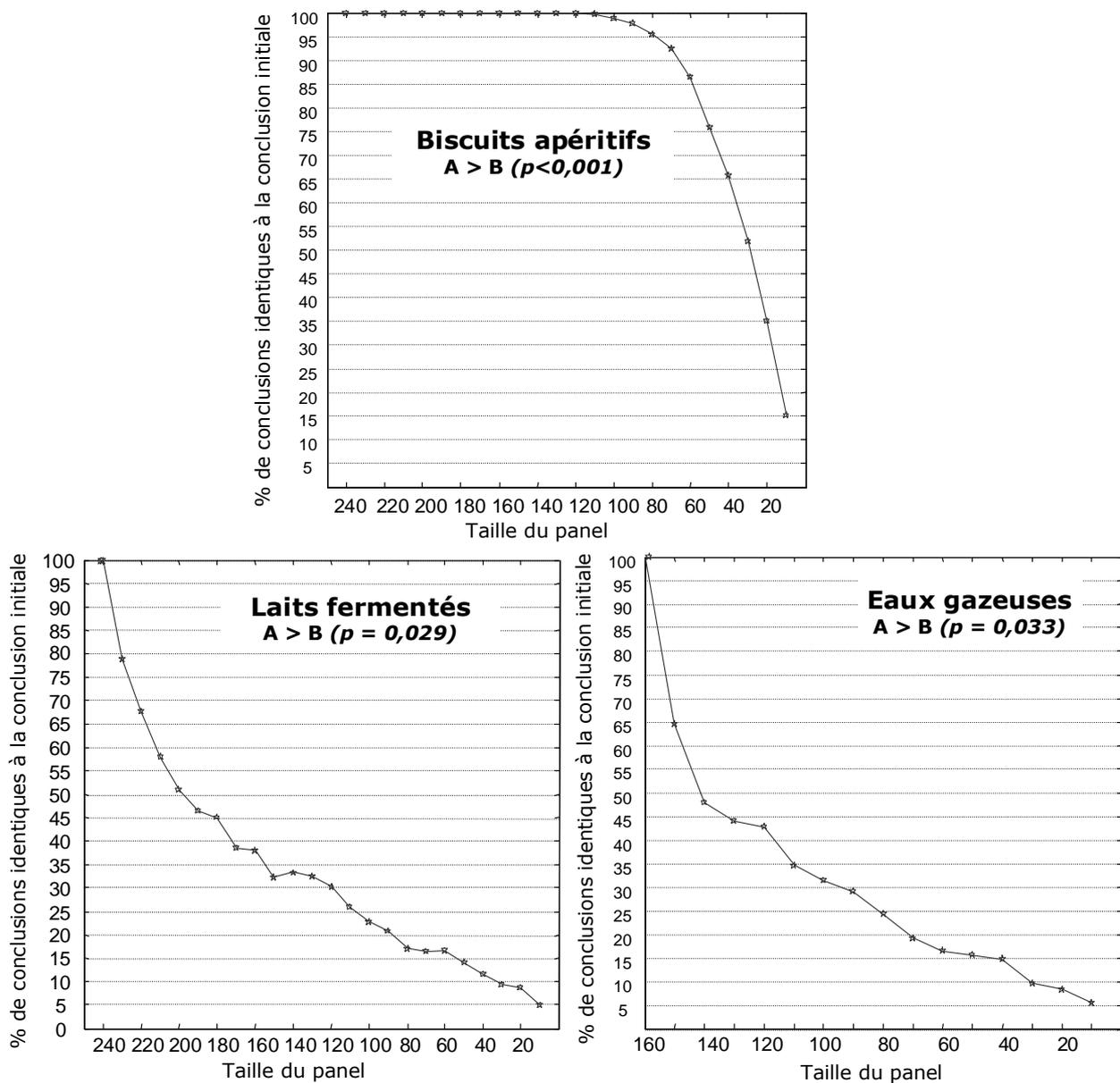


Figure 26 : Tests à domicile réalisés sur les trois catégories de produits : évolutions du pourcentage de panels aléatoires qui aboutissent à une différence significative d'appréciation entre les produits en fonction de l'effectif de sujets conservé

III - PRESENTATION DES PRODUITS : MONADIQUE PUR OU MONADIQUE SEQUENTIEL ?

Les trois expérimentations présentées dans l'article 1 (p. 69) nous ont permis de réaliser une comparaison de la présentation des produits en monadique pur et en monadique séquentiel.

Tout d'abord, nous avons observé que les notes hédoniques obtenues en monadique pur étaient souvent légèrement plus élevées que celles obtenues en monadique séquentiel, que celles-ci soient récoltées en salle ou à domicile. Ce résultat confirme l'effet de sur-notation du premier produit fréquemment relevé dans la littérature. Cependant, cet effet d'ordre dans chacune des expérimentations n'est significatif que pour deux jeux de données sur les six analysés (trois tests en salle et trois tests à domicile).

En revanche nous observons pour quatre jeux de données sur six, que l'analyse des données en monadique séquentiel aboutit à une différence significative d'appréciation alors que l'analyse des données en monadique pur ne la montre pas. Bien entendu, le fait que les conclusions des procédures en monadique séquentiel soient obtenues sur la base d'un nombre de données deux fois plus élevé que les conclusions des procédures en monadique pur influence positivement le pouvoir discriminant. Cette meilleure discrimination des procédures en monadique séquentiel par rapport aux procédures en monadique pur peut également provenir du fait que les données sont alors appariées et que l'analyse statistique fait intervenir l'effet *sujet* dans le modèle d'ANOVA qui permet de diminuer la résiduelle et d'aboutir plus facilement à un effet *produit* significatif. Cependant cette différence du niveau de discrimination peut également provenir tout simplement de la possibilité pour les sujets de comparer les deux produits testés en monadique séquentiel alors que le monadique pur ne le permet pas. Nous pouvons illustrer l'impact d'une évaluation comparative ou dans l'absolu sur les résultats en comparant les résultats obtenus sur le premier produit dégusté (évaluation dans l'absolu) avec ceux obtenus sur le deuxième produit dégusté (évaluation comparative). Par exemple, les résultats de l'étude sur les laits fermentés présentés dans le tableau 20 mettent clairement en évidence qu'à effectif équivalent, les produits sont plus discriminés quand ils sont évalués en deuxième position qu'en première position. Ce phénomène a été observé de la même manière pour l'ensemble des études réalisées en monadique séquentiel, que ce soit en salle ou à domicile. Ainsi, outre le nombre de données plus élevé, le meilleur pouvoir discriminant des procédures en monadique séquentiel par rapport aux procédures en

monadique pur proviendrait également de la possibilité pour le sujet de noter le deuxième produit en comparaison au premier produit.

Tableau 20 : Comparaison des résultats obtenus sur le premier produit dégusté et sur le deuxième produit dégusté pour l'étude des laits fermentés

		PRODUITS DEGUSTES EN PREMIER		PRODUITS DEGUSTES EN SECOND	
		Produit A	Produit B	Produit A	Produit B
Test en salle	Effectif	120	119	119	120
	Note moyenne	7,22	<i>NS</i> 7,15	6,70	** 5,96
	<i>SD</i>	1,80	1,84	1,95	2,24
Test à domicile	Effectif	120	121	121	120
	Note moyenne	7,17	<i>NS</i> 7,10	7,30	* 6,80
	<i>SD</i>	1,94	1,63	1,62	1,92

De manière plus fondamentale ou encore logistique, les deux procédures ont chacune leurs avantages présentés dans le tableau 21.

Le principal avantage de la procédure en monadique pur est sa capacité à récolter des évaluations hédoniques a priori plus proches de ce qu'expérimente un consommateur dans sa vie quotidienne lors de laquelle il consomme rarement deux versions proches d'un produit l'un à la suite de l'autre. En effet, de manière naturelle, le jugement de l'aliment se ferait en comparaison avec le produit consommé habituellement plutôt qu'avec un autre échantillon imposé. Ainsi, les procédures en monadique pur seraient particulièrement adéquates dans le cas de rénovation de formule ou de réduction de coût avec des sujets consommateurs réguliers du produit testé. Cependant, nous pouvons nous interroger sur la pertinence d'une évaluation en monadique pur quand le sujet n'a pas de référence interne (produit innovant). De la même façon, les procédures en monadique séquentiel dans le cas de problématiques de comparaison par rapport à la concurrence ont également du sens étant donné que le consommateur est confronté aux deux produits en même temps sur le marché.

Par ailleurs, Schwartz et Pratt (1956) ont observé qu'un intervalle de temps élevé entre des évaluations successives permet de diminuer l'effet d'ordre. Ainsi, alors que le monadique séquentiel mono-séance du test en salle mettrait clairement en place une évaluation comparative artificielle des deux produits, le test à domicile en monadique séquentiel multi-séance permettrait un mécanisme qui se rapproche peut être plus de celui mis en place lors

d'évaluations en monadique pur ou tout du moins se rapprochant d'un jugement plus naturel (c'est-à-dire une comparaison par rapport à la mémoire).

Tableau 21 : Les raisons de préférer une procédure de présentation des produits en monadique pur ou en monadique séquentiel

Les raisons de préférer le monadique pur	Les raisons de préférer le monadique séquentiel
- Jugement par rapport à une référence interne (produit habituellement consommé) et non par rapport à un autre échantillon testé.	- Plus rapide pour l'évaluation d'un grand nombre de produits (surtout à domicile).
- Possibilité, en salle, de consommer une dose normale (naturelle) de l'échantillon.	- Moins coûteux car pas besoin d'un groupe de sujets par produit testé.
- Constitution de banques de données envisageable.	- L'analyse des données prend en compte l'effet sujet.

Les différents arguments en faveur de l'une ou l'autre méthode rendent difficile la recommandation d'une procédure plutôt que l'autre et expliquent peut-être la pratique parfois observée de la double analyse des données en monadique séquentiel et en monadique pur sur le premier produit évalué (appelée « proto-monadique » par les instituts d'études). Dans le cas d'une faible différence d'appréciation détectée en monadique séquentiel, nous pouvons supposer que celle-ci aurait de grandes chances de ne pas être observée en proto-monadique. Même si ce genre de résultats incohérents complique grandement la prise de décision, cette double analyse pourrait être utilisée comme indicateur de confiance dans le résultat obtenu.

IV - LES QUESTIONS DIAGNOSTIQUES

IV.1 - DES INFORMATIONS HEDONIQUES SUPPLEMENTAIRES

Une fois l'ensemble des évaluations globales recueillies pour tous les produits, les tests mis en place et présentés dans ce document proposaient l'évaluation de quelques questions diagnostiques. Les résultats de l'analyse de ces questions diagnostiques ont été peu détaillés dans ce document bien que ceux-ci aient été généralement riches d'informations.

Les questions diagnostiques nous ont par exemple permis de différencier deux types de schémas lors de l'observation d'une différence non-significative d'appréciation globale :

- les échantillons ne sont pas non plus discriminés sur les questions diagnostiques,
- les échantillons sont discriminés sur les questions diagnostiques bien que leur appréciation soit identique.

Cette information est particulièrement intéressante dans le cadre de nos comparaisons de méthodes en salle et à domicile. Par exemple, alors que les deux méthodes aboutissent à des conclusions différentes concernant l'écart d'appréciation globale des deux crackers salés au fromage (significatif pour le test à domicile et non significatif pour le test en salle), les réponses aux questions diagnostiques sont en revanche équivalentes et discriminent bien les deux recettes selon les deux méthodes. Ainsi, la non significativité de l'écart d'appréciation des deux recettes en salle ne provient pas de l'impossibilité pour les sujets de différencier les produits.

Par ailleurs, la comparaison des tests en salle traditionnels et situationnels (tests en situations de repas et de petit déjeuner) a été enrichie par l'analyse des questions diagnostiques. Par exemple, nous avons observé que la situation de petit déjeuner modifie grandement les appréciations partielles (goût et forme) des petits pains croquants par rapport à un test en salle traditionnel alors que le test en situation de repas ne semble pas modifier la notation de quelques caractéristiques sensorielles des eaux gazeuses par rapport à un test en salle traditionnel. Nous avons également pu élaborer des hypothèses sur la contribution des caractéristiques sensorielles hédoniques des produits à la construction du jugement global qui semble différente selon les conditions d'évaluation.

IV.2 - COMMENT INTRODUIRE CES QUESTIONS DANS LES PROTOCOLES ?

En ce qui concerne la façon d'introduire ces questions diagnostiques dans les tests, la revue bibliographique concernant l'influence de l'évaluation d'attributs spécifiques sur les évaluations globales nous a montré la nécessité de séparer les deux types d'évaluations afin de ne pas biaiser les évaluations hédoniques globales récoltées. Ainsi, la mise en place de deux phases distinctes d'évaluation lors des tests en salle en monadique séquentiel semble un bon compromis permettant de satisfaire à la fois le besoin de récolter des évaluations globales pures et l'attrait pour le recueil d'informations diagnostiques, sans trop alourdir les protocoles de test. En ce qui concerne les tests à domicile, le délai qui sépare l'évaluation du premier produit de celle du second produit est supposé diminuer l'impact des évaluations spécifiques du premier produit sur l'évaluation globale du deuxième produit. Ainsi, la nécessité de séparer les évaluations est peut-être moins forte que dans le cas du test en salle. Cependant, nous pourrions également imaginer que, lors de la dernière visite de l'intervieweur au domicile du sujet pour récolter l'évaluation globale du deuxième produit, celui-ci lui propose de re-déguster et de ré-évaluer les deux produits sur la base des questions diagnostiques.

IV.3. QUELLES QUESTIONS DIAGNOSTIQUES ?

Demander directement au consommateur de faire le lien entre les caractéristiques sensorielles des produits et leurs préférences/rejets est souvent proposé dans les études consommateurs sous forme de questions ouvertes (qu'est ce qui vous plait/déplait dans ce produit ?). Cependant étant donné que la plupart des réactions affectives existent sans une analyse consciente des caractéristiques sensorielles, je ne crois pas en l'utilité de ces questions ouvertes qui aboutiront à des réponses probablement basées sur des croyances. De la même façon, les questions d'opinion proposant des échelles d'accord (type Likert) posées lors de quelques études menées dans ce travail se sont montrées peu informatives avec très souvent plus de 20% des réponses attribuées à la catégorie « sans opinion » et un pouvoir discriminant très faible. En revanche, l'analyse des questions de type JAR ou d'appréciations partielles concernant quelques caractéristiques spécifiques sensorielles des produits a confirmé la capacité des sujets à différencier les produits sur ce type de questions. Concernant plus spécifiquement les questions JAR, la comparaison des produits sur la base des réponses aux catégories « trop » ou « pas assez » était généralement plutôt cohérente avec la comparaison des profils sensoriels. Dans ce sens, je pense que les questions de type JAR ou appréciations partielles doivent être conservées dans les études pour identifier des pistes de compréhension de succès ou d'échec organoleptiques.

Il convient cependant d'être conscient des limites de l'analyse de ces questions qui ne doivent en aucun cas servir d'outils de description des échantillons. Aujourd'hui, le problème des questions diagnostiques réside dans le fait que les attributs sensoriels évalués découlent souvent de l'analyse descriptive qui nécessite un entraînement des sujets « experts » avant d'obtenir un consensus. La méthode de « *repertory grid* » décrite par Gains (1989) et Thomson et McEwan (1988) proposant à chaque consommateur de construire son propre vocabulaire pour différencier les échantillons avant de réaliser leur évaluation est une approche qui commence à être utilisée pour pallier ce problème (Veinand *et al.*, 2006).

V - EVALUATION PONCTUELLE ET DYNAMIQUE DES PREFERENCES

Même si ce travail n'aborde pas spécifiquement le problème de la dynamique des préférences, la comparaison des tests en salle et des tests à domicile présente intrinsèquement la comparaison de résultats basés sur une exposition ponctuelle (test en salle) ou répétée (test à domicile) des sujets avec les échantillons testés. Les expositions répétées lors du test à domicile sont supposées être en grande partie responsables de la meilleure validité prédictive

du test à domicile en comparaison au test en salle. En effet, la prédiction du niveau d'appréciation d'un produit sur le long terme ne serait pas envisageable par l'intermédiaire de la première impression du sujet concernant le produit. D'après Köster et Mojet (2006), le processus mis en jeu lors de la construction de la première impression est différent de celui mis en jeu lors d'un contact plus durable. La première impression ne serait pas concernée par le plaisir ressenti par le sujet mais uniquement par son souhait de réduire son incertitude concernant le produit et les données hédoniques récoltées lors des tests ponctuels seraient alors basées sur de la curiosité. Une fois les incertitudes du sujet résolues et sa curiosité rassasiée, un deuxième processus centré sur l'aspect hédonique peut alors se mettre en place lors de contacts ultérieurs avec le produit et aboutir à une opinion relativement stable.

Ainsi, les tests en salle basés sur une consommation ponctuelle des produits ne pourraient pas être révélateurs de l'appréciation du produit sur le long terme. Cependant, il semble compliqué de proposer plusieurs contacts avec le produit au sein d'un design de test en salle traditionnel. Parmi nos propositions méthodologiques, seul le test d'authenticité présente l'avantage d'imposer plusieurs expositions aux produits. De manière plus directe, le test de lassitude proposé par Köster (1998) consiste à faire évaluer plusieurs échantillons du même produit les uns à la suite des autres. Cependant, la difficulté de ce test réside dans le fait que les répétitions sont réalisées à l'insu du sujet. En effet, le sujet ne sait pas qu'il déguste en fait le même échantillon et peut donc dans un soucis de cohérence se forcer à donner des notes d'appréciations différentes bien qu'il apprécie tous les échantillons de la même manière.

Plus spécifiquement, l'évolution des préférences pose un problème pour les tests ponctuels réalisés sur des produits innovants lors desquels des réactions artificielles de rejet (néophobes) ou d'attrait (néophiles) risquent d'être recueillies. Proposer une phase d'exposition des sujets aux produits à domicile (ce qui est généralement possible en aveugle car les produits testés sont des produits prototypes réalisés en station pilote) suivie d'une phase d'évaluation des produits en salle permettrait de collecter un jugement hédonique construit sur plusieurs consommations du produit et obtenu dans des conditions de test contrôlées. De manière plus large, ce trait de personnalité concernant la néophobie des sujets soulève d'autres problèmes liés à la façon de réaliser les recrutements des sujets. En effet, d'après Lambert (1999), les consommateurs sont généralement néophobes et, lors de la réalisation de recrutement, ce sont plus généralement les sujets néophiles qui acceptent de participer aux tests de consommation (Meiselman *et al.*, 2006). Lors du congrès « A Sense of Diversity » (2006) Meiselman soulève donc le problème que nous sommes peut être en train de développer des produits pour les consommateurs les plus favorables et que nous oublions

alors les consommateurs moins tolérants envers la nouveauté. Cette caractéristique est également quelque chose auquel j'ai été confrontée lors des phases de recrutement des consommateurs en rue avec une difficulté à recruter les personnes plus âgées qui refusent de participer aux tests contrairement aux personnes plus jeunes. Etant donné que le marché des « seniors » s'impose, lentement mais inexorablement, comme un des segments émergents de consommation du XXI^{ème} siècle (les *papy boomers*), il devient urgent de trouver un moyen d'accéder facilement à l'opinion de cette nouvelle cible de consommateurs.

VI - QUESTIONS METHODOLOGIQUES RELATIVES AU TEST EN SALLE

VI.1 - LA PREPARATION DE L'ALIMENT TESTE

La majorité des produits testés par les entreprises agroalimentaires lors d'études consommateurs sont des aliments de grande consommation qui nécessitent généralement peu de préparation individuelle. Ainsi, je pense qu'il serait assez simple de proposer, de manière contrôlée, une consommation des produits préparée de manière individuelle. Aujourd'hui, Danone propose par exemple de questionner le sujet sur la température à laquelle il consomme habituellement une catégorie de boisson et de lui présenter la boisson à la température adéquate. De la même façon, si un consommateur est habitué à consommer un soda en cannette, pourquoi lui servir l'échantillon à tester dans un verre ? Proposer du sucre pour des yaourts semble également assez simple à réaliser et serait sans aucun doute d'un grand apport pour améliorer la prédiction des données hédoniques collectées en salle. Par ailleurs, le contrôle des conditions de consommation permettrait également d'analyser les données par type d'usage et éventuellement d'identifier la préparation qui permet de récolter une performance maximale de la recette.

Les données obtenues lors des tests situationnels illustrent l'intérêt de proposer une préparation individuelle des produits. En ce qui concerne par exemple les deux recettes de petit pain croquant testées lors du test en salle traditionnel, l'analyse des questions d'appréciations partielles posées en salle permet de souligner une faiblesse du produit A par rapport au produit B concernant son goût mais pas sa forme ronde. L'interprétation de ces données en termes actionnables pour les équipes de développement serait donc la nécessité de travailler sur l'amélioration du goût du produit A. Au contraire, l'ajout d'une composante de préparation individuelle lors de l'évaluation des produits (tartinables, boisson chaude, trempage) a mis en évidence que le goût du petit pain, une fois qu'il est tartiné et potentiellement trempé dans une boisson chaude est autant apprécié que le goût du produit B.

En revanche, la manipulation du produit met en évidence une faiblesse concernant la forme ronde du produit A moins appréciée que la forme ovale du produit B. Ainsi, l'amélioration du produit A devrait peut-être passer par la rénovation de sa forme plutôt que de son goût.

VI.2 - ETAT PHYSIOLOGIQUE DES PARTICIPANTS ET QUANTITE CONSOMMEE

L'examen de la littérature montre que la faible quantité consommée lors des tests en salle est un inconvénient majeur dans l'obtention de données prédictives de la réalité. La pratique des tests a été particulièrement révélatrice de ce problème. En effet, nous nous sommes rendu compte de la difficulté de faire consommer une dose minimale de produit par les consommateurs. L'étude sur les biscuits d'apéritif a par exemple mis en évidence la consommation d'un demi-biscuit par un grand nombre de participantes ! La consommation imposée d'une quantité minimale est difficile à mettre en œuvre et ne semble pas être particulièrement efficace. Bien évidemment, les procédures de test en monadique pur devraient permettre, plus facilement que les procédures en monadique séquentiel, la consommation d'une dose naturelle de consommation mais ce n'est pas systématiquement l'effet que l'on observe. Il reste donc à trouver des moyens indirects pour amener le sujet à consommer plus de produit sans qu'il se sente forcé. Les horaires de test semblent également intervenir sur la quantité consommée. Ainsi, les séances de dégustation réalisées en fin de matinée (avant le déjeuner) ou en milieu et en fin d'après-midi sont clairement plus favorables à des quantités consommées élevées. Il serait donc intéressant d'ajouter un critère de sélection basé sur le délai du dernier repas par exemple ou encore de ne pas interviewer les consommateurs en début d'après-midi. Cependant, la littérature a mis en évidence un manque de discrimination des sujets dans un état de faim élevé qui apprécient fortement tout ce qu'ils consomment. Ainsi, la difficulté de ce problème réside dans le fait que le niveau de faim ou de soif des sujets sélectionnés doit leur permettre de consommer une quantité de produit raisonnable tout en conservant leur esprit critique envers les produits. Il serait assez simple d'intégrer dans la phase de recrutement une question sur l'envie de consommer la catégorie de produits testée avec une échelle de réponse proposant des quantités de produit croissantes et de ne sélectionner alors que les sujets rapportant un état de faim ou de soif intermédiaire. Par ailleurs, il faudrait également préciser aux participants que peu de produits seront évalués afin d'être certain qu'ils ne se « réservent » pas parce qu'ils pensent qu'ils devront déguster de nombreux produits.

B – IMPACT DES CONDITIONS ARTIFICIELLES DE DEGUSTATION SUR LA MESURE HEDONIQUE

Dans la littérature, un grand nombre de travaux a mis en évidence un impact des conditions d'évaluation naturelles vs. artificielles sur les résultats. Cependant, cet effet n'étant pas systématiquement observé, nous avons souhaité l'évaluer dans le cadre des produits commercialisés par Danone. Dans cette optique, nous avons mené cinq expérimentations (salle vs. domicile) sur des catégories de produit différentes qui nous ont alors permis d'identifier et d'éclaircir plusieurs points concernant l'impact des conditions de test artificielles en salle sur le jugement hédonique.

I - DES NOTES HEDONIQUES EN SALLE MOINS ELEVEES QU'A DOMICILE

Ce phénomène de sous-notation du test en salle par rapport au test à domicile a été observé sur la majorité des jeux de données récoltés dans ce travail. Cette sous-notation des produits évalués en conditions artificielles de dégustation par rapport aux notes obtenues en situation naturelle de consommation a été également relevée dans la littérature. Plusieurs hypothèses ont été soulevées pour interpréter ce niveau de notation plus élevé à domicile : des conditions moins formelles que le test en salle, un environnement plus confortable, des conditions de consommation optimales, un effet de simple exposition dû à la consommation répétée des produits. Chacune de ces hypothèses nécessiterait d'être validée. L'état d'humeur ou plus exactement le bien-être global des sujets a été soulevé comme variable explicative de la différence du niveau de notation obtenue lors d'un test en salle et d'un test à domicile. En effet, l'appréciation globale des produits testés est certainement soumise à un effet d'assimilation dépendant de l'état d'humeur du sujet. Il serait donc intéressant, afin de tester cette hypothèse, d'évaluer l'état d'humeur global des sujets lors d'un test en salle et d'un test à domicile par le biais d'une échelle comme la PANAS (*Positive And Negative Affect Scale* : Watson *et al.*, 1988). De notre côté, nous avons exploré l'amélioration des conditions de consommation en salle dans le développement de nos différentes propositions méthodologiques (tests situationnels). Cependant, nous n'avons pas systématiquement observé un niveau de notation plus élevé que lors du test en salle traditionnel. Le test en situation de petit déjeuner ne met pas en évidence un niveau de notation supérieur à celui du test en salle traditionnel. Le test en situation de repas, quant à lui, met en évidence une augmentation du niveau de notation par rapport au test en salle traditionnel. Cependant, ce

niveau de notation est également supérieur à celui du test à domicile. L'impact de la rémunération sur l'état d'humeur des participants est probablement plus en cause dans cette sur-notation (par rapport aux tests traditionnels dans lesquels les sujets n'étaient pas rémunérés) que l'amélioration de l'à-propos des conditions de consommation. Ainsi, même si, d'après la bibliographie, rémunérer les participants permettrait de les impliquer plus fortement et diminuerait la manifestation du biais de « *satisficing* » dans les réponses, celle-ci induit également un manque d'esprit critique de la part des participants et des scores très élevés.

Ce phénomène de sous-notation en salle n'a pas été, au premier abord, particulièrement préoccupant étant donné que les tests organoleptiques sont mis en place pour comparer la performance de plusieurs recettes. Ainsi, si les écarts de notes restent les mêmes entre les produits, les conclusions statistiques des tests et in fine les décisions stratégiques peuvent ne pas être modifiées même si le test en salle met en évidence des notes plus basses que le test à domicile. Cependant, nous nous sommes finalement aperçus que certains travaux de développement de nouveaux produits ou d'extension de gamme dans l'entreprise utilisent des standards d'action basés, entre autres, sur la nécessité d'obtenir des notes d'appréciation supérieures à 7 (sur une échelle de 1 à 10). Bien entendu, le choix d'un niveau de note comme indicateur de rejet ou d'acceptation d'un produit est largement critiquable. Par ailleurs, le choix de cette note limite de 7 est loin d'être clairement justifié, voire même arbitraire. Quoiqu'il en soit, le décalage de notation en fonction des conditions d'évaluation dans le cadre de l'utilisation de ce type de standards d'action est embarrassant. En effet, compte tenu de nos résultats, il est évident qu'une note de 7 obtenue lors d'un test en salle ne représente pas le même niveau d'appréciation qu'une note de 7 obtenue lors d'un test à domicile. Si nous reprenons par exemple les notes moyennes pour les boissons à base de jus de fruit et de lait nous observons que, lors du test en salle, seule la recette standard obtient une note supérieure à 7 alors que, lors du test à domicile, les deux recettes à teneurs réduites en sucre obtiennent des scores d'appréciation supérieurs à 7. Ces résultats nécessitent donc de retravailler les standards d'actions ou, tout du moins, de les différencier en fonction des conditions de test.

Outre le problème engendré par les standards d'action basés sur un niveau de note, la variabilité du niveau de notation en fonction des conditions d'évaluation rend également l'utilisation des bases de données obtenues en monadique pur particulièrement risquée. En effet, nous assistons à l'heure actuelle à un désir des industriels de mettre en place leurs études consommateurs avec une présentation des produits en monadique pur. Ce design de protocole est particulièrement recherché puisqu'il permet de s'affranchir de l'influence des autres stimuli testés sur l'appréciation de chaque échantillon. Ces jugements hédoniques

construits dans l'absolu par des groupes de participants différents ont l'avantage de permettre une capitalisation des données de tests consommateurs aujourd'hui particulièrement recherchée par les industriels car elle permet d'économiser la mise en place d'un test hédonique à chaque nouveau besoin concernant un nouveau produit ou un nouvel échantillon à tester. Ainsi l'influence des conditions de consommation sur le niveau de notation que nous avons observé remet également en question la pertinence de la capitalisation des données dans un objectif de comparaison de notes hédoniques obtenues lors de différentes études. Nous pouvons par exemple supposer que les niveaux de notation de deux produits évalués lors de deux études réalisées en salle à des moments différents de l'année (été vs. hiver) peuvent être différents. Nous recommandons donc la plus grande prudence lors de la comparaison de notes d'appréciation données à des produits dans des environnements différents même si ces derniers ont été évalués en monadique pur. Ainsi, les industriels doivent mettre en place un système de traçabilité des conditions d'évaluation dans lesquelles les notes ont été obtenues afin de s'assurer de la possibilité de comparer deux notes obtenues lors d'études différentes.

II – DIFFICILE D'ANTICIPER L'IMPACT DES CONDITIONS ARTIFICIELLES DE DEGUSTATION SUR LES CONCLUSIONS STATISTIQUES DES TESTS

II.1. LES CONCLUSIONS STATISTIQUES NE SONT PAS SYSTEMATIQUEMENT DIFFERENTES DE CELLES OBTENUES LORS D'UN TEST A DOMICILE

Les cinq expérimentations menées nous ont permis d'identifier plusieurs caractéristiques, liées aux produits testés, qui doivent être prises en compte lors du choix de la mise en place d'un test en salle ou d'un test à domicile.

Les conditions de consommation habituelles de la catégorie de produits influencent très certainement la possibilité d'obtenir des résultats différents entre un test en salle et un test à domicile. Ainsi, un test en salle réalisé avec des produits habituellement consommés lors d'occasions très différentes les unes des autres (ex : eaux gazeuses) ou lors d'une occasion de consommation rapide et individuelle (ex : laits fermentés) a de faibles probabilités d'aboutir à une conclusion différente de celle d'un test à domicile. En revanche, un test en salle réalisé avec un produit consommé généralement lors d'occasions bien spécifiques et en combinaison d'autres aliments (ex : biscuits feuilletés au fromage) risque fortement d'aboutir à une conclusion hédonique différente de celle d'un test à domicile.

La nature des différences sensorielles entre les échantillons comparés peut également intervenir dans l'augmentation de la probabilité d'obtenir des résultats différents entre un test

en salle et un test à domicile. Ainsi, un test en salle pour comparer des produits différents sur une caractéristique sensorielle comme le sucré (salé, matière grasse ?) a de fortes probabilités d'entraîner des résultats différents de ceux d'un test à domicile.

Ces différentes caractéristiques nous ont donc permis de mettre en place une grille d'hypothèses (tableaux 9 et 10, p 109-110) à la disponibilité des utilisateurs de tests consommateurs pour les aider à anticiper les risques qu'ils prennent lorsqu'ils décident de mettre en place un test en salle à la place d'un test à domicile en fonction des produits testés.

II.2 - VARIABILITE DE LA NATURE DE L'INFLUENCE DES CONDITIONS D'EVALUATION SUR LES CONCLUSIONS STATISTIQUES DES TESTS

Cette première partie expérimentale a mis en évidence que l'influence du contexte sur les conclusions des tests reste particulièrement complexe et très difficile à anticiper.

Contrairement à ce qui est décrit dans la littérature, nous n'avons jamais observé de classements de produits inversés selon la méthode employée mais cela ne nous permet en aucun cas de conclure que le test en salle permet de récolter un classement hédonique des produits testés identique à celui du test à domicile. En effet, nos expérimentations ont été menées avec seulement deux ou trois produits ce qui rend des inversions de classement moins probables que dans le cas de produits plus nombreux.

Alors que nous pouvons faire l'hypothèse que des notes hédoniques obtenues en salle seront systématiquement moins élevées que des notes hédoniques obtenues à domicile et que le classement hédonique ne sera pas entièrement modifié, il nous est en revanche beaucoup plus difficile d'anticiper l'évolution de la discrimination des échantillons testés. En effet, même si plusieurs de nos expérimentations ont mis en évidence un impact des conditions d'évaluation sur les conclusions statistiques des tests, la nature de cet impact n'est pas stable. Alors que le test en salle sur les biscuits d'apéritif est moins discriminant que le test à domicile, celui réalisé sur les boissons à base de jus de fruit et de lait est plus discriminant que le test à domicile. Par ailleurs, la variabilité de la nature de l'impact des conditions d'évaluation a été également observée au sein d'une seule et même étude (thé glacé) dans laquelle les hommes sont plus discriminants en salle et les femmes à domicile.

Cette impossibilité de conclure qu'une méthodologie est plus discriminante qu'une autre vient bousculer l'idée préconçue d'un meilleur pouvoir discriminant du test en salle apporté par la présentation des produits en monadique séquentiel mono-séance (par opposition à la présentation en monadique séquentiel multi-séance du test à domicile).

C - DISCUSSION SUR LES CRITERES DE QUALITE D'UNE METHODE DE TEST CONSOMMATEURS

Dans l'article 5 (annexe I), nous amorçons une discussion concernant la pertinence des critères de qualité généralement utilisés pour valider une méthode de test consommateurs. Les méthodes de mesure hédonique en analyse sensorielle étant historiquement dérivées de l'analyse sensorielle descriptive, les expérimentateurs ayant tenté de juger de la qualité de leurs données se sont basés sur les critères habituellement utilisés dans le domaine de la métrologie. Ainsi, nous retrouvons dans les tentatives de validation des méthodes de test consommateurs les principaux critères de validation des méthodes de mesure. Cependant, l'adaptation de ces critères métrologiques à la mesure hédonique est quelque peu abusive si nous considérons les nombreux mécanismes psychologiques à la base de la formation du jugement hédonique des consommateurs. Il convient de reprendre la signification de ces critères de validation dans le domaine des tests consommateurs à la lumière des erreurs d'interprétation des données consommateurs citées par Köster (2003) ainsi que sur la base des données obtenues dans ce travail.

I - LE CRITERE D'EXACTITUDE DE LA METHODE

I.1 – DISCUSSION SUR L'UTILISATION DU TEST A DOMICILE COMME METHODE DE REFERENCE

L'utilisation de ce critère d'exactitude pour valider une méthode suppose la détermination d'une méthode de référence. Dans ce travail, nous avons fait le choix d'utiliser les données obtenues lors des tests à domicile comme données de référence. Ce choix a été remis en cause plusieurs fois tout au long de la thèse, que ce soit de manière théorique ou par les résultats que nous avons obtenus.

En effet, le test à domicile, tel qu'il est pratiqué actuellement, présente certains inconvénients comme par exemple le manque de contrôle des conditions de test. Les praticiens s'interrogent même sur le fait que certains sujets ne consomment pas les produits distribués ou rapportent l'opinion de l'ensemble du foyer plutôt que leur opinion personnelle. En ce qui concerne les données hédoniques que nous avons obtenues dans notre travail, les résultats de la question d'intention de consommation des deux crackers salés au fromage selon les situations de consommation ont également généré des doutes quant à la qualité des

données hédoniques obtenues lors du test à domicile. En effet, en opposition au test à domicile, ce test a mis en évidence une supériorité de la recette B par rapport à la recette A. Selon notre hypothèse de travail, ce résultat devrait nous pousser à rejeter la proposition méthodologique basée sur l'amorçage de différentes situations de consommation. Cependant, nous nous devons également de remettre en question la validité de la conclusion hédonique obtenue par le test à domicile tel qu'il a été mis en place. Dans un ouvrage récent consacré aux tests consommateurs, Köster et Mojet (2006) proposent cinq règles qui, d'après eux, sont nécessaires pour faire du test à domicile un bon candidat dans la prédiction de l'appréciation d'un produit sur le long terme (figure 27).

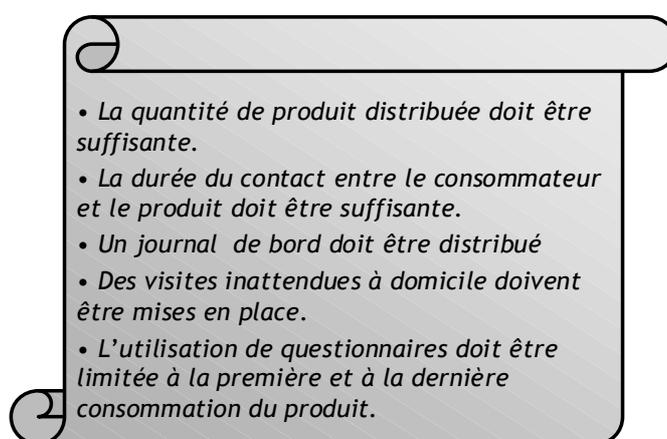


Figure 27 : Les 5 commandements de Köster et Mojet (2006) pour mettre en place un bon test à domicile

Ainsi, d'après ces cinq règles qui semblent tenir du bon sens,

- La quantité de produit distribuée doit être ajustée à la taille de la consommation normale de la famille et doit permettre au moins sept répétitions de consommation par toute la famille.
- La durée du contact entre le consommateur et le produit doit être suffisante pour permettre au sujet d'utiliser le produit à une fréquence normale de consommation. Ceci rend donc le test à domicile non approprié pour des produits dont la fréquence de consommation est très faible (moins de deux fois par mois).
- Un journal de bord doit être fourni dans lequel la personne interrogée note les quantités utilisées du produit et les quantités d'autres produits de la même catégorie achetés et consommés pendant la durée du test.
- Des visites inattendues à domicile doivent être mises en place afin de poser des questions générales sur la préparation du produit, la facilité d'utilisation de l'emballage et pour vérifier que le produit et le journal sont bien utilisés.

- L'utilisation de questionnaires doit être limitée à la première et à la dernière fois que la famille consomme le produit durant l'expérience. Le premier questionnaire doit consister uniquement en une note hédonique du produit et des questions factuelles sur l'expérience passée concernant ce type de produit. Le second questionnaire doit demander en premier une note hédonique puis évaluer une série de questions sensorielles sur des échelles « *Just About Right* ».

Bien évidemment, peu de protocoles de tests à domicile mis en place en routine respectent l'ensemble de ces règles compte tenu du coût que cela engendrerait.

I.2 – PROPOSITION DE METHODES DE REFERENCE BASEES SUR DES DONNEES COMPORTEMENTALES

Si nous nous intéressons au fondement du jugement hédonique et à sa mesure par le biais de questionnement, l'erreur fondamentale est de croire que les gens font des choix rationnels (Tversky *et al.*, 1990). En fait, lorsque nous devons justifier de nos choix, de nos actions, nous essayons de les rationaliser et en faisant cela nous les falsifions. Ceci se traduit dans les méthodes de mesure déclaratives par l'impossibilité pour une personne de ne pas donner de réponse à une question, même si celle-ci est en principe impossible à fournir. Les gens sont polis et ne veulent pas décevoir l'enquêteur. Ils ne veulent pas non plus paraître stupides et sont donc capables de donner une réponse à n'importe quelle question. Les réponses données sont donc plus souvent le reflet de croyances ou de désirabilité sociale que d'un réel jugement. Les gens aiment manger et boire sans se soucier d'analyser leurs expériences alimentaires et, s'ils le font occasionnellement, ils ont beaucoup de difficultés à s'exprimer sur ces expériences. Ainsi, quand ils sont confrontés à une simple question à propos de leur aliment préféré, ils doivent inventer une réponse pour satisfaire l'enquêteur. Dans un tel contexte, nous comprenons donc que les mesures hédoniques déclaratives ne soient pas utilisées comme mesure de référence. Les mesures comportementales en revanche permettraient de vérifier que les consommateurs font réellement ce qu'ils déclarent faire.

Dans l'idéal, les données du marché pourraient être l'indicateur de référence de l'appréciation réelle d'un aliment par les consommateurs. Cependant, ces données de vente ne sont pas uniquement dépendantes des qualités organoleptiques de l'aliment en question car de nombreuses variables économiques et marketing influencent l'achat d'un aliment par rapport à un autre. Ainsi, ce n'est pas parce que les données hédoniques obtenues lors d'un test consommateurs en aveugle ne reflètent pas exactement les données de ventes sur le marché que la méthode de test mise en place pour obtenir l'appréciation des produits doit être remise

en cause. Une solution afin de rendre les données du marché pertinentes pour valider une méthode de test hédonique en aveugle serait, par exemple, l'utilisation des données de vente de différents produits d'une même gamme ayant bénéficié de la même stratégie marketing (marque, prix, promotion, distribution) et de comparer ces données de vente avec les données hédoniques en aveugle de ces différentes versions du produit. Bien que cette approche soit séduisante, elle reste compliquée à mettre en œuvre étant donné que les différentes versions d'un même produit sont rarement lancées en même temps sur le marché.

L'examen de la littérature concernant les méthodes utilisées pour accéder à l'appréciation d'aliments par les consommateurs a mis en évidence l'existence de méthodes de mesure comportementales à l'aveugle qui mériteraient d'être utilisées comme méthode de référence. Il s'agit par exemple de l'utilisation de distributeurs automatiques de produits gratuits dans des lieux publics de consommation. Vickers *et al.* (1999) étudient la quantité consommée de différentes versions de laits en mettant à la disposition des consommateurs d'une cafétéria pendant plusieurs semaines une machine qui distribue alternativement les différentes versions de lait. Hoyer *et al.* (2003) proposent le même type de procédure pour améliorer le pouvoir prédictif des mesures hédoniques. Les produits codés sont dégustés préalablement par les consommateurs en laboratoire qui repartent avec une carte magnétique qui leur permettra de choisir dans un distributeur gratuit pendant plusieurs semaines, sur leur lieu de travail, les produits qu'ils ont envie de consommer. Cette méthode permet de récolter par le biais de la carte magnétique les différents choix individuels qui ont été faits pendant toute la période du test. Cette approche comportementale basée sur le choix des produits dans des situations naturelles de consommation est également en train de se développer par l'intermédiaire de cafétérias expérimentales. Ainsi plusieurs projets de partenariat entre des instituts de recherche sur le comportement alimentaire et des entreprises de restauration d'entreprise sont en train de se construire afin de permettre l'étude du comportement alimentaire dans des environnements naturels tout en permettant un contrôle des conditions de consommation (le « Restaurant of the futur » à l'Université de Wageningen ou encore le projet « Cafet'n lab » financé par l'INRA dans le cadre du Programme de Recherche en Alimentation). L'utilisation de ces cafétérias expérimentales sera d'un grand apport pour collecter des données hédoniques comportementales valides.

II - LE CRITERE DE REPETABILITE DES SUJETS

La répétabilité de la mesure est l'aptitude d'une méthode à donner, dans des conditions d'utilisation définies, des réponses très voisines lors de l'application du même signal d'entrée. Dans le cadre des méthodes de test consommateurs, nous pourrions donc définir la répétabilité au niveau de la stabilité de la mesure donnée par un même panel de sujets. Nous nous contentons donc ici d'étudier la répétabilité au niveau du groupe de sujets (c'est-à-dire l'instrument de mesure) et non au niveau individuel. En effet, croire que la préférence d'un consommateur est stable est une idée fautive. Les changements de préférence individuels ont largement été démontrés autant sur le long terme (Moskowitz, 2000 ; Sulmont, 2000) que sur le court terme (Leon *et al.*, 1999 ; Levy, 1998 ; Pliner, 1982 ; Porcherot et Issanchou, 1998). L'opinion du consommateur change réellement car celui-ci a une mémoire et sa seconde expérience avec un aliment sera forcément influencée par sa première expérience. De la même façon, la théorie de « *preference reversal* » de Tversky *et al.* (1990) considère que les individus ne font pas de choix rationnel et que leurs préférences sont donc constamment en mouvement.

Que ce soit au niveau individuel ou au niveau du groupe de consommateurs, la répétabilité de la mesure d'appréciation des aliments par les sujets n'a jamais été étudiée spécifiquement dans ce travail. Comme nous l'avons précisé précédemment, nous sommes bien évidemment conscients de l'instabilité des préférences individuelles. Cependant, le test en situation de repas en monadique séquentiel sur les eaux gazeuses fournit un jeu de données sur l'évaluation répétée de l'appréciation des différents composants des deux repas (taboulé, tarte et salade de fruit). Je n'ai donc pas pu résister à l'exploration de la répétabilité des sujets sur ces notes hédoniques données aux composants des repas lors des deux séances espacées d'une semaine. L'ANOVA (*note hédonique = sujet + produit + séance + produit × séance*) montre que l'effet *séance* [$F_{(1,795)} = 1,41$; $p = 0,235$] ainsi que l'interaction *produit × séance* [$F_{(2,795)} = 0,73$; $p = 0,481$] ne sont pas significatives, indiquant une relativement bonne répétabilité du groupe sur les deux séances quel que soit le produit. Par ailleurs, l'effet *produit* significatif [$F_{(2,795)} = 307,91$; $p < 0,001$] illustre que les trois produits sont bien discriminés en termes d'appréciation ce qui confirme la capacité des sujets à être répétables tout en étant discriminant. Cependant, l'instabilité des préférences des consommateurs ayant été mise en évidence lors de nombreuses études dans la littérature, nous ne pouvons pas exiger d'une méthode qu'elle donne des résultats répétables pour le même groupe de consommateurs interrogés.

Alors que le critère de répétabilité considère des répétitions de la mesure dans un court délai et dans les mêmes conditions (validation intra-laboratoire), le critère de reproductibilité considère des répétitions de la mesure en modifiant légèrement certaines conditions (validation inter-laboratoire). La généralisation des données hédoniques collectées pour un groupe de sujets à une population plus globale (validité externe) nécessite une stabilité de la conclusion obtenue sur un autre panel de sujets appartenant à la même population. Ainsi, ce critère se rapprocherait du critère de reproductibilité avec une modification de l'appareil de mesure (le groupe de consommateurs). Cette capacité de reproductibilité des méthodes traditionnelles en salle et à domicile ou encore des méthodes proposées dans cette thèse n'a pas été étudiée. Ce travail reste donc à réaliser dans une perspective de validation des solutions méthodologiques proposées.

III - LE POUVOIR DISCRIMINANT DE LA METHODE

En métrologie, le pouvoir discriminant de la méthode est son aptitude à permettre de faire la distinction de manière significative entre des stimuli proches. Il est cependant important de se poser la question de l'utilité d'un pouvoir discriminant maximum dans le cadre des évaluations hédoniques. En effet, l'objectif de ces tests n'est pas de mettre en évidence une différence entre les échantillons mais de révéler et éventuellement comparer l'appréciation des produits par les consommateurs. Deux produits peuvent tout à fait être perçus différemment tout en étant appréciés de la même façon (c'est le cas des eaux gazeuses en situation de repas). Ainsi, nous ne pouvons pas décentement exiger d'une méthode de test hédonique qu'elle soit la plus discriminante possible. Forcer la discrimination hédonique des produits pourrait entraîner de mauvaises décisions stratégiques notamment dans le cadre de problématiques de réduction de coût dans lesquelles les standards d'action sont basés sur une parité entre la nouvelle et l'ancienne formule. Ainsi, une méthode « trop » discriminante pourrait entraîner l'annulation du lancement d'une nouvelle formule moins coûteuse pour l'entreprise et aurait donc des conséquences économiques importantes.

Finalement, la seule caractéristique indispensable concernant le pouvoir discriminant d'une méthode concerne sa nécessité d'être représentatif de la discrimination hédonique des produits lors d'un acte naturel de consommation. Malgré tout, ce critère de discrimination maximum des échantillons évalués en termes d'appréciation reste proposé pour juger de la qualité d'une mesure hédonique et est encore présent dans certaines études de comparaison de méthodes hédoniques (Barylko-Pikielna *et al.*, 2004).

IV - LE CRITERE DE ROBUSTESSE

La robustesse d'une méthode est sa capacité à limiter les variations de résultats lorsqu'elle est soumise à des variations liées à l'environnement, aux procédures, au laboratoire, au personnel, etc. Compte tenu de l'influence des variables de contexte précédemment discutées, nous comprenons donc bien qu'il est illusoire de croire en la stabilité d'une mesure hédonique obtenue dans des conditions de test différentes. C'est d'ailleurs dans ce souci de robustesse des données hédoniques que les expérimentateurs se sont longtemps employés à contrôler et à fixer les variables de contexte dans la mise en place de leurs tests en conditions standardisées.

Outre la robustesse des données face à la perturbation de l'environnement de dégustation, nous avons, de notre côté, exploré ce critère de robustesse face à un autre type de perturbation souvent négligée lors de la mise en place des tests consommateurs : la modification de la taille du groupe de consommateurs interrogé.

Lors de l'obtention d'une différence significative d'appréciation, ce critère de robustesse de la conclusion hédonique obtenue face à la diminution de l'effectif du panel s'est montré pertinent afin de tempérer les conclusions de l'étude si celle-ci est facilement déstabilisée (article 5 en annexe I). Nous pourrions utiliser ce critère comme outils de « validation interne » de la conclusion hédonique obtenue c'est-à-dire comme indicateur du degré de certitude de la conclusion obtenue. Cependant, ce critère de robustesse nécessite encore un travail de mise au point afin de pouvoir tester la robustesse d'une conclusion qui aboutit à une différence non significative entre les produits. Ce travail nécessiterait la détermination délicate d'un écart de notation seuil en dessous duquel l'appréciation des deux produits pourrait être considérée comme équivalente. Les nombreux jeux de données en possession des instituts d'études ainsi que leur expérience seraient d'une aide considérable pour cette tâche. Par exemple, Sauerhoff *et al.* (2005) proposent d'utiliser un écart de 0,5 points sur une échelle hédonique à 9 points.

D - PROPOSITIONS METHODOLOGIQUES : APPLICATION EN ROUTINE ?

Les applications des protocoles proposés dans ce travail nous ont permis d'évaluer leur potentiel concernant leur facilité de mise en place et la richesse des informations hédoniques récoltées. Je souhaiterais maintenant reconsidérer chacun des protocoles proposés dans le cadre de leur application dans un contexte industriel routinier.

I – LES APPROCHES ORIENTEES « SITUATION »

I.1 - LES TESTS SITUATIONNELS : DES METHODOLOGIES A SIMPLIFIER

Les protocoles de test situationnel ont été mis en place de manière à ce que le statut d'un grand nombre de variables contextuelles soit approprié à la situation de consommation mimée. Ce souhait de proposer des conditions de consommation les plus appropriées possible nous a conduit à mettre en place des protocoles relativement complexes et lourds. Dans une perspective d'application de ces protocoles en routine, le test en situation de petit déjeuner semble abordable avec des contraintes de coût et de logistique raisonnables. Il semble en revanche plus compliqué de mettre en œuvre les protocoles du test en situation de repas et du test dans un environnement social tel qu'ils ont été appliqués dans ce travail. En fait, un moyen de simplifier ces protocoles de test situationnel serait d'explorer des conditions de dégustation en salle de plus en plus appropriées en essayant d'identifier les variables qui nécessitent réellement d'être dans un statut naturel et celles dont le statut artificiel standardisé n'a finalement peut-être pas besoin d'être modifié. King *et al.* (2004) ont, de cette façon, observé que proposer la consommation d'un repas entier (entrée, plat principal et boisson) n'a pas d'impact sur l'appréciation globale du plat principal. Ainsi, les aliments consommés en plat principal pourraient ne pas nécessiter systématiquement l'environnement alimentaire naturel d'un repas complet pour être évalués de manière pertinente. En revanche, les produits d'accompagnement sont sensibles au fait d'être dégustés seul ou en accompagnement d'un repas.

I.2 - PROTOCOLE A PRIVILEGIER EN FONCTION DE L'UNIVERS DE CONSOMMATION DU PRODUIT

La mise en place de ces tests situationnels suppose l'hypothèse d'une consommation mono-situationnelle du produit. Cependant, de nombreux produits sont consommés de manière multi-situationnelle et le fait d'imposer un seul contexte de consommation ne permet alors pas de révéler les performances des produits dans d'autres situations d'utilisation tout aussi importantes. L'analyse des informations d'usage de l'univers produit est donc nécessaire avant de faire le choix d'utiliser ce type de protocole de test en situation. En effet, les petits pains croquants sont effectivement consommés dans la majorité des cas lors d'un petit déjeuner ce qui justifie donc de proposer des conditions de consommation proches d'un petit déjeuner. En revanche, les données d'usage de la catégorie des eaux gazeuses ont mis en évidence que celles-ci sont consommées en situation de repas uniquement pour la moitié des occasions de consommation. De fait, un tel protocole en situation de repas ne nous permet pas d'accéder de manière globale à l'appréciation des produits selon la diversité des conditions de consommation habituelles. Même s'il est indispensable de connaître les performances de ces eaux gazeuses en situation de repas, les autres situations de consommation peuvent tout à fait entraîner des résultats différents du test en situation de repas. Nous avons par exemple observé que, lors du test à domicile, les consommatrices révélaient une différence d'appréciation significative entre les deux marques d'eau gazeuse ($A > B$). En revanche, le test en situation de repas ne mettait pas cette différence d'appréciation en évidence. Ainsi, nous pouvons imaginer que la différence d'appréciation obtenue à domicile pour les deux marques provient des conditions de consommation en dehors des repas pendant lesquelles le produit A est probablement plus apprécié que le produit B. Le test en salle traditionnel qui proposait une dégustation des deux eaux gazeuses dans des conditions standardisées a d'ailleurs abouti à cette différence d'appréciation.

La mise en place de plusieurs tests en conditions appropriées en fonction des différentes situations de consommation étant impossible dans un contexte industriel, cela nous a poussé à explorer d'autres moyens pour accéder à cette information hédonique dépendante des situations de consommation. Le protocole de test basé sur l'utilisation de la mémoire épisodique du participant en aidant le sujet à se projeter dans ses propres situations quotidiennes a été développé dans ce sens.

La figure 28 résume le type de protocoles de test que nous proposons en fonction de l'univers de consommation des produits testés.

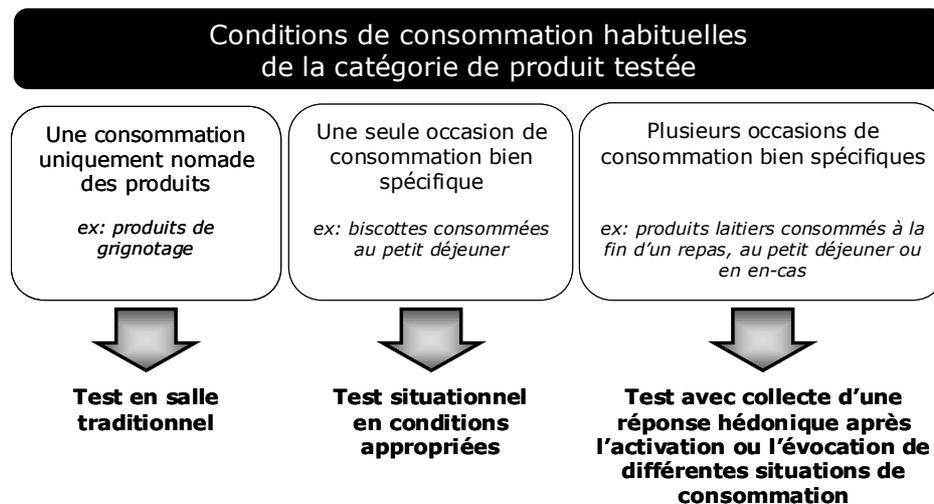


Figure 28 : Protocoles de test à privilégier en fonction de l'usage des produits

I.3 - PROPOSITION DE DIFFERENTS PROTOCOLES BASES SUR L'UTILISATION DE LA MEMOIRE EPISODIQUE DU PARTICIPANT

Dans le cas de produits pouvant être consommés selon plusieurs situations bien identifiées, nous proposons donc l'application du protocole de test basé sur l'utilisation de la mémoire épisodique du participant qui permet de récolter deux types d'informations :

- une appréciation globale émise après l'amorçage chez le participant d'un épisode de consommation,
- une intention de consommation en fonction de différentes situations évoquées par le biais de scénarios auditifs.

La collecte de ces deux types de données a été testée de manière combinée dans le cadre de ce travail bien qu'elles ne répondent pas tout à fait à la même problématique.

L'amorçage d'un épisode de consommation a pour objectif d'aider le participant à activer en mémoire les conditions habituelles de consommation du type de produit testé. Cette procédure ne permet pas réellement de récolter une mesure hédonique en fonction des différentes conditions de consommation mais propose de récolter une information hédonique globale construite sur la base des signaux situationnels habituellement activés lors d'un acte naturel de consommation.

L'intention de consommation récoltée en fonction de différentes situations évoquées permet, quant à elle, de prendre en compte cet aspect multi-situationnel de manière plus approfondie et a pour objectif de déterminer si les performances des produits comparés sont identiques selon les situations de consommation qui véhiculent des attentes différentes.

Dans le cadre traditionnel de la comparaison des performances organoleptiques de plusieurs produits, la combinaison de ces deux types de protocoles (de mesures) n'est pas forcément indispensable.

L'application de la mesure d'intention d'utilisation en fonction des situations est intéressante à partir du moment où elle permet d'identifier clairement des interactions *produit* \times *situation*. En effet, si un produit semble plus apprécié qu'un autre sur l'ensemble des situations, l'appréciation globale après amorçage d'un épisode de consommation devrait permettre d'accéder à cette information. L'étude de produits proches sensoriellement risque probablement de ne pas mettre en évidence de différence d'appréciation en fonction des situations. Cette proximité sensorielle entre les produits est une des hypothèses que nous avons soulevé pour expliquer le fait que les deux recettes de crackers présentent une interaction *produit* \times *situation* non significative. En revanche, des produits présentant des caractéristiques sensorielles plus différentes ont probablement plus de chance de mettre en évidence des classements hédoniques différents selon les situations générant des attentes sensorielles différentes.

Dans le cadre de la comparaison de produits proches sensoriellement (problématiques de réduction de coût ou de rénovation de produits), nous recommandons donc plutôt l'usage uniquement de l'appréciation globale obtenue l'après activation autobiographique d'un épisode de consommation.

En revanche, les problématiques de comparaison des performances d'un produit par rapport à ses concurrents, les phases préliminaires de développement de produits ou encore les problématiques d'extension de gamme (ex : différents parfums) seraient probablement enrichies par la collecte combinée d'une information hédonique globale traditionnelle et d'une information hédonique en fonction des différentes situations d'usage habituelles de la catégorie de produits. La nature de la question hédonique posée en fonction des situations mériterait d'être cependant approfondie. Poser une question de préférence envers les différents produits pour chaque situation (Henry, communication personnelle) pourrait par exemple être envisagé. Par ailleurs, il conviendrait également de trouver un moyen de valider les scénarios utilisés dans un tel protocole afin d'être certain de leur efficacité pour évoquer la situation d'intérêt.

II – LES APPROCHES ORIENTEES « IMPLICATION »

II.1 - QUEL EST LE TEST QUI PERMET L'IMPLICATION LA PLUS FORTE DES PARTICIPANTS ?

La sensibilité des jugements hédoniques aux conditions de dégustation nous a poussé à proposer, lors de tests en salle, des moyens d'engager plus fortement le participant dans ses réponses dans l'espoir de diminuer l'impact négatif des conditions de consommation artificielles. Bien entendu, cette voie d'amélioration basée sur l'implication du participant dans ses réponses n'est pas incompatible avec les tests précédemment évoqués. En effet, nous nous sommes attachés, dans cette partie exploratoire, à proposer des mesures hédoniques différentes de la simple question d'appréciation ou de préférence. Nous pourrions donc tout à fait envisager la mise en place d'un test de choix ou d'authenticité dans les conditions d'évaluation d'un test situationnel ou encore après l'activation chez le participant d'un épisode de consommation de la catégorie de produits testée.

En ce qui concerne l'hypothèse de travail basée sur le test à domicile comme méthode de référence, les deux protocoles de test appliqués aux biscuits feuilletés au fromage ont permis d'aboutir avec succès à une conclusion identique à celle du test à domicile c'est-à-dire une préférence significative pour la recette A. Le tableau 22 présente les résultats de la question de préférence obtenus à domicile, à la fin du test d'authenticité et pour la proposition 1 du test de choix expérimentaux.

Tableau 22 : Résultats de la question de préférence entre les deux biscuits feuilletés obtenus selon les protocoles de test.

	Test à domicile	Test d'authenticité	Test de choix expérimentaux
Effectif	240	217	236
% de sujets qui préfèrent le biscuit A	65,5%	57,6%	60,2%
p-value (test binomial)	p < 0,001	p = 0,015	p < 0,001

Nous remarquons que le test de choix expérimentaux permet de conclure plus significativement que le biscuit A est préféré au biscuit B par rapport au test d'authenticité. Ainsi, le test de choix permet de mieux s'approcher de l'intensité de la préférence révélée par le test à domicile. Cette plus forte discrimination du test de choix sur le test d'authenticité

pourrait être attribuée à une plus forte implication des participants dans le test de choix que dans le test d'authenticité.

Lors du test de choix et du test d'authenticité, nous avons proposé à la moitié des sujets interviewés de répondre à un questionnaire permettant d'évaluer leur niveau de motivation dans la tâche expérimentale qui leur a été demandée (l'autre moitié des sujets a été interrogée sur d'autres aspects du protocole). Ce questionnaire est basé sur une traduction de l'échelle IMI (Intrinsic Motivation Inventory) qui a déjà été utilisée pour évaluer l'expérience subjective de participants par rapport à une activité réalisée lors d'expérimentations (Deci *et al.*, 1994 ; Ryan, 1982). Cette échelle présente initialement l'évaluation de 7 dimensions différentes (45 items) qui sont toutes reliées à la motivation du participant pour réaliser la tâche. Nous avons fait le choix de nous intéresser à cinq de ces dimensions (effort/importance, intérêt/plaisir, stress/pression, valeur/utilité perçue, compétence perçue) et d'utiliser trois items pour chacune de ces dimensions. Le tableau 23 présente un exemple d'item pour chacune des dimensions évaluées dans le questionnaire. A la fin du test, les participantes devaient donc répondre au questionnaire en précisant à quel point elles étaient d'accord avec chaque item proposé en utilisant une échelle en 7 points allant de « pas du tout d'accord » (1) à « totalement d'accord » (7).

Tableau 23 : Exemples d'items pour chaque dimension évaluée par le questionnaire

DIMENSION	EXEMPLE D'ITEM
Effort/importance	<i>Cela me tenait à cœur de réaliser ce test correctement</i>
Intérêt/plaisir	<i>J'ai trouvé ce test ennuyeux</i>
Stress/pression	<i>J'étais détendue en réalisant ce test</i>
Valeur/utilité perçue	<i>J'ai l'impression d'avoir été utile en participant à ce test</i>
Compétence perçue	<i>J'ai le sentiment de ne pas avoir été capable de faire correctement ce test</i>

Nous avons donc comparé les données récoltées par le biais de ce questionnaire lors du test de choix expérimentaux (116 sujets) et lors du test d'authenticité (122 sujets). Ainsi, le tableau 24 présente les scores moyens obtenus pour chaque dimension selon la tâche expérimentale effectuée par les sujets : le test d'authenticité ou la tâche de choix expérimentaux. Bien entendu, ces résultats sont à prendre avec précaution dans la mesure où les propriétés psychométriques du questionnaire ont été peu étudiées. Par ailleurs, les données obtenues pour les deux types de tests ont été récoltées pour deux groupes de participantes différents, ce qui rend difficile la distinction entre l'influence de la tâche et l'influence du

groupe de sujets sur les scores. Cependant, ces données concernant la façon dont les sujets ont expérimenté les deux types de tests aboutissent à plusieurs informations qu'il convient de mentionner.

Tableau 24 : Notes moyennes et écart-types obtenus pour chaque dimension selon le protocole de test et résultats des tests de Student

	Test d'authenticité	Test de choix expérimentaux	t de Student	p-value
Effort/importance	5,37 (1,54)	5,61 (1,76)	1,88	0,060
Intérêt/plaisir	5,11 (1,69)	5,51 (1,71)	3,13	0,002
Stress/pression	1,91 (1,29)	1,62 (1,33)	2,89	0,004
Valeur/utilité perçue	4,04 (1,97)	4,47 (2,00)	2,94	0,004
Compétence perçue	4,16 (1,87)	4,93 (2,11)	5,09	< 0,001

La tâche d'authenticité semble avoir été vécue plus difficilement par les sujets que la tâche de choix expérimentaux. En effet, les sujets se sont sentis plus compétents, plus satisfaits de leur performance dans la tâche de choix que dans la tâche d'authenticité. En ce qui concerne le plaisir à réaliser la tâche et son intérêt, le test de choix semble également plus performant que la tâche d'authenticité. Contrairement à ce que nous aurions pu imaginer, les participantes se sont également senties plus utiles en participant au test de choix qu'en participant au test d'authenticité qui présente pourtant un enjeu plus important et des conséquences sur du plus long terme que le test de choix. L'effort fourni par les participantes pour réaliser correctement la tâche d'authenticité serait également moins important que l'effort fourni pour réaliser la tâche de choix expérimentaux. Enfin nous observons que les participantes étaient plus nerveuses pendant la tâche d'authenticité que pendant la tâche de choix expérimentaux. Cette dimension, contrairement aux quatre autres, serait un prédicteur négatif de la motivation intrinsèque. Nous retrouvons peut-être ici le facteur de stress lié aux conditions d'enquêtes propice à la manifestation du *satisficing* (Esposito et Jobe, 1991). L'amorçage négatif préalable à la tâche d'authenticité est probablement en cause dans la manifestation du stress chez les participantes inquiètes de ne pas réaliser correctement le test. Il serait particulièrement intéressant de tester la possibilité d'amorcer émotionnellement les sujets de manière positive plutôt que négative afin de supprimer ce stress.

L'analyse globale de l'ensemble des réponses aux 15 items par ACP (résultats non détaillés) met en évidence une dimension regroupant 7 items qui pourrait être attribuée à la mesure de l'implication du participant dans le test. La note moyenne obtenue sur cette

dimension pour le test d'authenticité (moyenne = 4,52 ; SD = 1,84) est significativement ($t = 4,98$, $p < 0.001$) moins élevée que la note moyenne obtenue pour le test de choix expérimentaux (moyenne = 4,99 ; SD = 1,90) ce qui confirmerait donc une implication générale moins importante des participantes dans la tâche d'authenticité que dans la tâche de choix expérimentaux. La mise en évidence de cette dimension d'implication dans notre questionnaire de motivation intrinsèque est tout à fait prometteuse. En effet, nous espérons pouvoir valider cet outil et l'utiliser afin d'évaluer la qualité d'un protocole de test consommateurs. Il serait ainsi intéressant de faire évaluer cette échelle d'implication par des sujets dans le cadre des autres protocoles mis en œuvre dans ce travail (test en salle, test à domicile, test situationnel et test avec activation de la mémoire épisodique du participant).

II.2 - PROTOCOLE ORIENTE « IMPLICATION » A PRIVILEGIER SELON LE CAS D'ETUDE

Comme dans le cas des propositions méthodologiques orientées « situation », la mise en œuvre pratique du test de choix ou du test d'authenticité suppose quelques conditions spécifiques qui en font des méthodes à appliquer dans le cas de problématiques différentes.

II.2.1 - Le test d'authenticité

Le protocole du test d'authenticité a été appliqué pour la comparaison de deux produits. Sa mise en œuvre lors de la comparaison de plus de deux produits reste possible mais entraînerait la dégustation d'un nombre d'échantillons important si nous souhaitons conserver l'analyse des données basée sur la théorie de la détection du signal qui suppose un minimum de répétitions de chaque produit.

Par ailleurs, si les produits sont très différents visuellement, l'utilité des répétitions reste à prouver étant donné que les sujets, une fois qu'ils auront identifié un des produits comme étant « authentique », répondront probablement de manière cohérente à toutes les répétitions. Dans le cas de la comparaison de deux produits très différents visuellement, nous pourrions conserver l'implication du participant par l'amorçage affectif négatif puis poser uniquement (sans répétition) une question sur l'identification des deux produits comme étant authentiques ou non. Cependant, cette simplification du protocole reste à tester et diminuerait grandement la richesse de l'information récoltée.

Ainsi, je pense que le protocole du test d'authenticité que nous proposons dans ce travail doit être utilisé dans le cas de produits proches sensoriellement comme lors de problématiques de réduction de coût et de rénovation de produit. Par ailleurs, ce type de protocole serait particulièrement approprié dans le cadre de l'étude de la fidélité des

consommateurs envers une recette. Enfin, les problématiques d'innovation de rupture pourraient également être enrichies par de telles données d'authenticité qui permettraient de vérifier l'adéquation entre un concept et différentes recettes proposées.

II.2.2 - Le test de choix expérimentaux non-hypothétiques

Comme nous l'avons précisé dans l'article concernant le protocole de choix expérimentaux, nous conseillons l'utilisation d'une séquence de troc basée sur la réduction de la quantité du produit préféré après une vraie dotation du lot du produit préféré. Encore une fois, ce protocole a été développé dans le cadre de la comparaison des performances de deux produits. Ainsi, l'adaptation de ce protocole de test à la comparaison de plus de deux produits reste à réaliser. Cette adaptation nécessiterait une organisation identique aux tests de comparaisons par paires multiples avec un plan de présentation en blocs incomplets dans le cas d'un grand nombre de produits. Ainsi, le participant serait confronté à plusieurs séquences de trocs pour différentes paires de produits et il tirerait au sort à la fin de la séance la proposition effective indiquant le lot de produit avec lequel il repartirait.

Cette méthode de choix expérimentaux suppose la possibilité de distribuer un grand nombre de produits ce qui complique son utilisation dans le cas de problématiques de réduction de coût, de rénovation ou d'extension de gamme dans la mesure où cela nécessite la production d'un grand nombre de produits pilotes. En revanche, cette méthodologie semble tout à fait adaptée pour comparer les performances organoleptiques de produits concurrents disponibles sur le marché.

Par ailleurs, la nature des produits testés intervient également dans la possibilité de mettre en place une telle méthodologie. Alors qu'il semble pertinent de distribuer au sujet une quantité importante de denrées non périssables à la fin de la séance, il serait nécessaire de mettre en place, pour les produits ultra-frais, un système coûteux de distribution à domicile des produits « gagnés » par le sujet. Une autre solution pour appliquer cette méthode avec des produits frais serait de réaliser le même type de troc basé sur un échange entre des bons d'achat de différentes valeurs pour l'un ou l'autre produit. Ce protocole de choix expérimentaux reste donc encore à optimiser pour être applicable à l'ensemble des catégories de produits. Pour l'instant nous préconisons donc plutôt l'utilisation du test d'authenticité dans le cas de problématiques mettant en œuvre des produits ultra-frais. L'utilisation du test de choix expérimentaux serait en revanche plus adéquate pour comparer les performances de produits non périssables.

II.2.3 - Synthèse

La figure 29 synthétise les cas pour lesquels nous pensons que les deux protocoles de test tels qu'ils sont présentés dans ce travail peuvent être appliqués selon :

- la nature des produits (périssables ou non),
- et l'écart sensoriel entre les produits comparés.

Nous observons donc que ni l'un ni l'autre de ces protocoles ne peuvent répondre à une problématique de comparaison de produits périssables très différents.

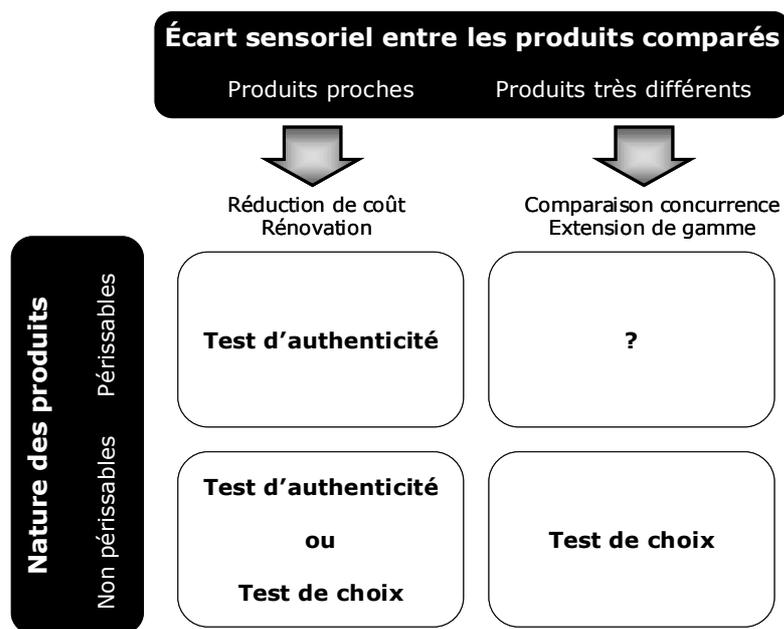


Figure 29 : Cadres d'application possible du test d'authenticité et du test de choix expérimentaux

Enfin nous nous devons également de mentionner que l'utilisation de ces deux types de protocoles suppose la comparaison d'au moins deux recettes différentes et qu'ils ne peuvent donc pas être mis en place dans le cas de l'évaluation de l'appréciation d'une seule recette (et in fine avec une procédure de présentation en monadique pur). Bien que ces deux protocoles développés dans la partie exploratoire de ce travail ont trouvé un accueil chaleureux de principe, leur utilisation en routine par les industriels est encore freinée par le désir de capitalisation des données hédoniques qu'ils ne permettent pas. Ainsi, convaincre les utilisateurs des tests consommateurs chez Danone du bénéfice de ces protocoles est aujourd'hui la délicate mission qu'il reste à accomplir.

CONCLUSION GENERALE

Ce travail de recherche nous a permis de mettre en évidence que les résultats obtenus en conditions d'évaluation artificielles ne traduisent pas systématiquement les performances des produits dans des conditions plus réalistes de consommation. Ce premier constat vient appuyer la remise en question de la pertinence du test en salle relevée dans la littérature. L'utilisation massive de ces tests en conditions standardisées et artificielles de consommation par les industriels de l'agroalimentaire pour mettre en place leurs stratégies de développement et de lancement de produit nous laisse imaginer l'ampleur des conséquences économiques. Par exemple, dans notre cas, lors du projet de développement concernant la réduction de la teneur en sucre de la boisson à base de jus de fruit et de lait, le test en salle aurait entraîné la décision de poursuivre les coûteux travaux de développement alors que le test à domicile met clairement en évidence que la formule testée avec 25% de sucre ajouté en moins est un bon candidat pour remplacer la formule actuelle. L'étude réalisée sur les pains suédois illustre également les conséquences économiques engendrées par des décisions prises sur la base de mauvais indicateurs d'appréciation obtenus par un test en conditions artificielles de consommation. En effet, en choisissant de pénétrer le marché des pains suédois, Danone s'est heurté à un concurrent particulièrement bien implanté. Les développeurs ont alors choisi de se différencier de ce principal concurrent en proposant un petit pain rond alors que la forme du *leader* est ovale. Cette décision a été prise sur la base de données hédoniques collectées lors d'un test en salle traditionnel dans lequel cette forme ronde n'avait pas été perçue comme moins appréciée que la forme ovale. Nous avons cependant constaté qu'en situation d'usage cette forme ronde s'est finalement montrée particulièrement peu pratique et pourrait expliquer la difficulté commerciale à laquelle est confrontée ce produit.

L'examen de la littérature a montré que de nombreux auteurs critiquent les conditions artificielles d'évaluation ainsi que la nature de la mesure hédonique collectée traditionnellement en sensoriel. Cependant, ces critiques à l'encontre de la mesure hédonique traditionnelle sont plus souvent le fruit de réflexions basées sur du bon sens que sur de réelles expérimentations. Le manque de données expérimentales concernant la mauvaise validité externe des tests en salle dans la littérature provient probablement d'une difficulté à laquelle nous avons également été confrontés dans ce travail c'est-à-dire l'incapacité de déterminer le bon/vrai résultat. Dans la littérature, cette difficulté à définir une méthode hédonique de référence se traduit par l'utilisation désordonnée de plusieurs mesures de références pour évaluer la qualité des méthodes de test en salle (quantité consommée ad libitum, choix d'un produit à emporter, appréciation obtenue sur le long terme,...). Par ailleurs, dans la majorité des études qui comparent des méthodologies de tests hédoniques, nous avons observé que les

auteurs se limitent souvent à une comparaison sur la base des résultats statistiques obtenus et principalement sur l'appréciation globale récoltée. Ce travail de thèse a soulevé la frustration de comparer les méthodes sur la seule base de cette mesure d'appréciation globale. Ainsi, je pense que nous avons tout intérêt à approfondir les façons de comparer les méthodes et à trouver d'autres indicateurs de qualité que ceux utilisés aujourd'hui. La comparaison des tests en salle traditionnels et situationnels a par exemple montré l'utilité de comparer les méthodes sur la base des questions diagnostiques qui peuvent apporter une information supplémentaire. L'appréciation globale des produits dégustés lors du test en situation de petit-déjeuner est par exemple équivalente à celle obtenue dans le test en salle classique alors que les appréciations partielles sont clairement influencées par les conditions d'évaluation. Mon expérience des terrains m'a également permis de constater des différences nettes concernant la façon dont les sujets perçoivent les différents types de tests auxquels ils participent. Ainsi, peut-être devrions nous explorer de nouveaux indicateurs de qualité basés sur cette perception du test par les sujets. Nous nous sommes intéressés à l'implication du participant dans le test mais d'autres critères pourraient être envisagés comme son état d'humeur, le temps qu'il passe dans la salle ou encore la quantité d'aliment qu'il consomme pendant le test.

En accord avec les nombreuses critiques faites aux méthodes de test traditionnelles, nous assistons donc dans le domaine de l'analyse sensorielle et des études consommateurs à un appel quasi-unanime des chercheurs pour trouver des méthodes plus fiables d'évaluation hédonique. Cependant, alors que les chercheurs se posent de nombreuses questions sur le type d'échelle hédonique à proposer ou encore sur la meilleure façon d'analyser les données consommateurs, nous avons finalement trouvé peu de propositions de protocoles de mesure radicalement différents de ceux utilisés traditionnellement. La mesure hédonique est un des domaines d'application initiaux de l'analyse sensorielle et date d'une cinquantaine d'années (Meiselman et Schutz, 2003). Cependant, les possibilités de recherche offertes par l'analyse sensorielle descriptive et ses outils de mesure ont assez rapidement entraîné les chercheurs à se désintéresser de la recherche méthodologique concernant la mesure hédonique. Ainsi, pendant plusieurs décennies, la mesure hédonique était présente dans les congrès d'analyse sensorielle uniquement dans le cadre d'études appliquées. Cependant, l'utilisation de l'analyse sensorielle descriptive par les industriels s'est rapidement révélée insuffisante pour prendre des décisions stratégiques de développement de produit et l'intérêt pour la mesure hédonique sensorielle a donc refait surface. Etant donné l'importance de cette mesure hédonique pour les industriels dans leur stratégie de développement, nous assistons donc, depuis une dizaine d'années, à un regain des recherches méthodologiques concernant cette

discipline. Ainsi, même si ce domaine de recherche est relativement jeune, il est aujourd'hui au centre des préoccupations puisqu'il concerne la moitié des travaux présentés lors des congrès d'analyse sensorielle comme le congrès Pangborn. L'apport de la psychologie à ce domaine d'étude s'est rapidement avéré nécessaire et nous avons alors assisté à de nombreux efforts afin d'introduire des concepts comme la distinction entre le comportement exploratoire et l'appréciation sur le long terme ou encore le rôle de la complexité des produits dans leur appréciation. Cependant, l'intégration de ces théories dans la façon de mener les études consommateurs reste exceptionnelle dans la recherche méthodologique et même inexistante dans la plupart des études menées en entreprise. Par ailleurs, aujourd'hui encore, l'utilisation de la psychologie dans la recherche sur le comportement alimentaire est souvent basée sur la théorie de l'action raisonnée (Fishbein et Ajzen, 1975) qui suppose que les gens font des choix rationnels et qu'ils peuvent expliquer les raisons à la base de leur comportement alors que, depuis une trentaine d'années, la recherche en psychologie a largement évolué et a soulevé le fait que les comportements ne sont finalement pas toujours sous contrôle conscient. Cette composante clé du comportement est aujourd'hui peu prise en compte dans la recherche méthodologique sensorielle qui continue à proposer des mesures actitudinales à travers des questionnements directs.

Alors que la recherche en méthodologie de test consommateurs se restreint à améliorer les outils de mesure existants principalement basés sur du questionnement direct, nous devrions maintenant nous intéresser aux méthodes pratiquées dans d'autres disciplines qui s'intéressent de plus près au comportement. Cette approche pluridisciplinaire est en train d'être amorcée comme en témoignent les interventions de plus en plus fréquentes de chercheurs en sociologie, en ethnologie ou encore en économie expérimentale qui sont invités à venir présenter leur façon d'appréhender le comportement des consommateurs lors des congrès d'analyse sensorielle. Par ailleurs, les programmes de recherche en France et à l'étranger concernant la construction de cafétérias expérimentales pour étudier les choix alimentaires en situation réelle de consommation sont également encourageants et ont l'immense avantage de regrouper des unités de recherche appartenant à différents domaines d'études du comportement des consommateurs (économie, gestion, analyse sensorielle, marketing...).

Dans ce travail de thèse, nous avons pu illustrer la richesse de l'information hédonique récoltée par l'intermédiaire de mesures hédoniques indirectes (l'intention de consommation dans une situation, le degré d'authenticité perçue, la propension à échanger) en utilisant des techniques empruntées à d'autres domaines d'études comme les scénarios utilisés en

sociologie, l'amorçage pratiqué en psychologie expérimentale ou les marchés expérimentaux provenant de l'économie expérimentale. Cette dernière approche s'est révélée particulièrement pertinente et je pense que les chercheurs ainsi que les industriels devraient se pencher sérieusement sur la façon d'introduire ces approches d'économie expérimentale dans leurs études consommateurs. Bien entendu, d'autres pistes de mesures hédoniques restent à étudier. Les mesures d'observation seraient par exemple une voie à explorer et plus particulièrement dans le cadre de la mesure des émotions. En effet alors que ces méthodes de mesure non-verbales étaient peu utilisées compte tenu de la difficulté de leur mise en place et du recueil fastidieux des données, le développement d'outils informatiques facilite aujourd'hui leur utilisation. Ainsi, les mimiques, les gestuelles ou encore les temps de réaction (de réponse) des sujets pourraient être de précieux indicateurs d'appréciation ou de certitude des sujets dans leur réponse. La recherche marketing peut également être riche de propositions méthodologiques notamment en ce qui concerne la façon d'appréhender les attentes et la satisfaction des consommateurs. En effet, les frontières entre les professionnels du marketing et ceux de l'analyse sensorielle établies principalement par des formations d'origines différentes (sciences humaines *vs.* sciences « dures ») ont contribué à nous faire perdre de vue que l'objectif de ces deux domaines d'études est identique. Le rapprochement de ces deux cellules de travail dans certaines entreprises est en ce sens plutôt encourageant.

Il semble donc que de nombreuses portes restent à ouvrir afin de découvrir de nouvelles façons de mener les études hédoniques pour accéder à l'opinion du consommateur. Le travail réalisé lors de cette thèse a cependant mis en évidence la difficulté d'adapter les méthodes relevées dans la littérature à une application industrielle. Ainsi, alors que les méthodes proposées sont souvent testées avec des espaces produits caricaturaux qui mettent en jeu des produits de catégories différentes, la recherche méthodologique en test consommateurs devrait se pencher plus souvent sur l'apport de ces méthodes dans le cadre des problématiques rencontrées en entreprise c'est-à-dire la comparaison de produits plutôt proches sensoriellement. Par ailleurs, plutôt que de nous étonner et de condamner les entreprises qui ne mettent pas en œuvre les recommandations méthodologiques que les chercheurs proposent, il devient également indispensable, lors du développement de nouvelles méthodes de garder à l'esprit leur application industrielle avec toutes les contraintes budgétaires et logistiques que cela suppose.

REFERENCES BIBLIOGRAPHIQUES

- AFNOR (2000). Norme XP V09-500 en cours de révision. Directives générales pour la réalisation d'épreuves hédoniques en laboratoire d'évaluation sensorielle ou en salle en conditions contrôlées impliquant des consommateurs. In. *AFNOR, recueil de normes, Analyse sensorielle*, 6ème édition.
- Alfnes F. (2004). Stated preferences for imported and hormone-treated beef: an application of the mixed logit model. *European Review of Agricultural Economics*, vol 31, p. 19-37.
- Alfnes F., Guttormsen A.G., Steine G., Kolstad K. (2007). Consumer willingness to pay for the color of salmon: a choice experiment with real economic incentives. *American Journal of Agricultural Economics*, In press.
- Allport G.W. (1935). Attitudes. In. Murchison C. (Ed.) *A handbook of social psychology* (p. 798-844). Worcester, MA : Clark University Press.
- Anderson E.W. (1998). Customer satisfaction and word of mouth. *Journal of Service Research*, vol 1, p. 5-17.
- Andrews F.M., Withey S.B. (1976). *Social indicators of well-being: Americans' perception of life quality*. New York : Plenum.
- Barylko-Pikielna N., Matuszewska I., Jeruszka M., Kozłowska K., Brzozowska A., Roszkowski W. (2004). Discriminability and appropriateness of category scaling versus ranking methods to study sensory preferences in elderly. *Food Quality and Preference*, vol 15, p. 167-175.
- Bayton J.A., Thomas C.M. (1954). Comparative single stimulus methods determining taste preferences. *Journal of Applied Psychology*, vol 38, p. 443-445.
- Beatty W.W. (1982). Dietary variety stimulates appetite in females but not in males. *Bulletin of the Psychonomic Society*, vol 19, p. 212-214.
- Becker G.M., deGroot M.H., Marschak J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, vol 9, p. 226-232.
- Bell J.A. (1992). The relationship of hedonics to behavior. Limitations and future of acceptance testing. *Activities Report of the RED Associates*, vol 44, p. 138-150.
- Bell R. (1993). Some unresolved issues of control in consumer tests: the effects of expected monetary reward and hunger. *Journal of Sensory Studies*, vol 8, p. 329-340.
- Bell R., Meiselman H.L. (1995). The role of eating environments in determining food choice. In. Marshall D.W. (Ed.) *Food choice and the consumer* (p. 292-310). Glasgow : Blackie Academic and Professional.
- Bellisle F., Giachetti I., Tournier A. (1988). Determining the preferred level of saltiness in a food product: a comparison between sensory evaluation and actual consumption tests. *Sciences des Aliments*, vol 8, p. 557-564.
- Berlyne D.E. (1960). *Conflict, arousal and curiosity*. New York: McGraw-Hill.

- Berlyne D.E., Slater J. (1957). Perceptual curiosity, exploratory behavior, and maze learning. *Journal of Comparative and Physiological Psychology*, vol 50, p. 228-232.
- Bernard P., Laporte V., Rodrigues E. (1993). Is there a relation between declared preferences and actual behaviour? *Viandes et Produits Carnes*, vol 14, p. 53-57.
- Berry S.L., Beatty W.W., Klesges R.C. (1985). Sensory and social influences on ice cream consumption by males and females in a laboratory setting. *Appetite*, vol 6, p. 41-45.
- Birch L.L., Billman J.B., Richards S.S. (1984). Time of day influences food acceptability. *Appetite*, vol 5, p. 109-116.
- Bitner M.J. (1990). Evaluating service encounters: the effects of physical surroundings and employee responses. *Journal of Marketing*, vol 54, p. 69-82.
- Bonin D. (2003). Variabilité évaluative et effets de contexte. Contribution à l'étude des limites de l'influence contextuelle implicite sur les réponses évaluatives et perspectives pour l'optimisation des situations de test de consommation. Thèse, Université Blaise Pascal, Clermont-Ferrand II, France.
- Bonin D., Chambres P., Bernard P. (2001). Influence du contexte sur la perception des produits alimentaires. In. Urdapilleta I., Ton Nu C., Saint-Denis C. (Eds.) *Traité d'évaluation sensorielle* (p. 223-232). Paris : Dunod.
- Boulze D., Montastruc P., Cabanac M. (1983). Water intake, pleasure and water temperature in humans. *Physiology and Behavior*, vol 97-102.
- Boutrolle I., Clément C., Sieffermann, J.M. (2003). Scales used in satisfaction survey questionnaires: a review. In. *IUFoST XIIIth World Congress of Food Science and Technology*, Chicago, IL.
- Boutrolle I., Arranz D., Rogeaux M., Delarue J. (2005). Comparing central location test and home use test results: Application of a new criterion. *Food Quality and Preference*, vol 16, p. 704-713.
- Boutrolle I., Delarue J., Arranz D., Rogeaux M., Köster E.P. (2007). Central location test vs. home use test: contrasting results depending on product type. *Food Quality and Preference*, vol 18, p. 490-499.
- Bower G.H. (1981). Mood and memory. *The American Psychologist*, vol 36, p. 129-148.
- Bower G.H., Gilligan S.G., Monteiro K.P. (1981). Selectivity of learning caused by affective states. *Journal of Experimental Psychology: General*, vol 110, p. 451-473.
- Brinberg D., McGrath J.E. (1985). *Validity and the research process*. Beverley Hills, CA : Sage
- Brunstrom J.M., Macrae A.W., Qannari E.M., Dijksterhuis G.B., Hunter E.A., MacFie H. (1997). Mouth state: A nuisance variable in preference tests? *Food Quality and Preference*, vol 8, p. 349-352.

- Buck D. (2003). Solicited commentary to Garber et al: measuring consumer response to food products. *Food Quality and Preference*, vol 14, p. 37-38.
- Burton M.D., Rigby D., Young T., James S. (2001). Consumer attitudes to genetically modified organisms in food in the UK. *European Review of Agricultural Economics*, vol 28, p. 479-498.
- Buzby J.C., Fox J.A., Ready R.C., Crutchfield S.R. (1998). Measuring consumer benefits of food safety risk reductions. *Journal of Agricultural and Applied Economics*, vol 10, p. 69-82.
- Cabanac M. (1971). Physiological role of pleasure. *Science*, vol 173, p. 1103-1107.
- Cabanac M. (1995). Palatability vs. money: experimental study of a conflict of motivations. *Appetite*, vol 25, p. 43-49.
- Cabanac M., Minaire Y., Adair E. (1968). Influence of internal factors on the pleasantness of a gustative sweet sensation. *Behavioral Biology*, vol 1, p. 77-82.
- Cabanac M., Pruvost M., Fantino M. (1973). Alliesthésie négative pour les stimuli sucrés après diverses ingestions de glucose. *Physiology and Behavior*, vol 31, p. 249-253.
- Cabanis R.F. (2003). The effect of asking current consumption questions before or after product evaluation. *Journal of Targeting, Measurement and Analysis for Marketing*, vol 12, p. 185-190.
- Calvin L.D., Sather L.A. (1959). A comparison of student preferences panels with a household consumer panel. *Food Technology*, vol 13, p. 469-472.
- Camerer C. (1995). Individual decision making. In: Kagel J.H., Roth A.E. (Eds.) *The handbook of experimental economics*. Princeton : Princeton University Press.
- Cardello A.V. (1994a). Consumer expectations and their role in food acceptance. In: MacFie H., Thomson D.M.H. (Eds.) *Measurement of food preferences* (p. 253-297). London : Blackie Academic and Professional.
- Cardello A.V. (1994b). Downsizing and other real-life issues. *Cereal Foods World*, vol 39, p. 184-187.
- Cardello A.V. (1995). Food quality: relativity, context and consumer expectations. *Food Quality and Preference*, vol 6, p. 163-170.
- Cardello A.V., Maller O. (1982). Acceptability of water, selected beverages and foods as a function of serving temperature. *Journal of Food Science*, vol 47, p. 1549-1552.
- Cardello A.V., Sawyer F.M. (1992). Effects of disconfirmed consumer expectations on food acceptability. *Journal of Sensory Studies*, vol 7, p. 253-277.
- Cardello A.V., Schutz H.G. (1996). Food appropriateness measures as an adjunct to consumer preference. *Food Quality and Preference*, vol 7, p. 239-249.

- Cardello A.V., Bell R., Kramer F.M. (1996). Attitudes of consumers toward military and other institutional foods. *Food Quality and Preference*, vol 7, p. 7-20.
- Cardello A.V., Schutz H.G., Snow C., Leshner L.L. (1996). The roles of food quality, preference, appropriateness and expectations on consumer satisfaction and consumption in a meal situation. In. *1996 IFT annual meeting: book of abstracts*.
- Cardello A.V., Schutz H.G., Snow C., Leshner L.L. (2000). Predictors of food acceptance, consumption and satisfaction in specific eating situations. *Food Quality and Preference*, vol 11, p. 201-216.
- Carifio J., Lanza M. (1992). The need to use control vignettes in social science research. *Work*, vol 2, p. 28-38.
- Chang A., Wilson M. (2004). Recalling emotional experiences affects performances on reasoning problems. *Evolution and Human Behavior*, vol 25, p. 267-276.
- Chiva M. (1992). Les aspects psychologiques des conduites alimentaires. In. Dupin H., Cup J.L., Malewiak M.I. (Eds.) *Alimentation et nutrition humaine* (p. 417-442). Paris : ESF.
- Chung S.J., Vickers Z.M. (2007). Influence of sweetness on the specific sensory satiety and long-term acceptability of tea. *Food Quality and Preference*, vol 18, p. 256-264.
- Clendenen V.I., Herman C.P., Polivy J. (1994). Social facilitation of eating among friends and strangers. *Appetite*, vol 23, p. 1-13.
- Cliff M.A., King M.C., Scaman C., Edwards B.J. (1997). Evaluation of R-Indices for preference testing of apple Juice. *Food Quality and Preference*, vol 8, p. 241-246.
- Creyer E.H., Ross W.T. (1997). The influence of firm behavior on purchase intention: do consumers really care about business ethics? *Journal of Consumer Marketing*, vol 14, p. 421-432.
- Dailliant B., Issanchou S. (1991). Most preferred level of sugar: rapid measure and consumption test. *Journal of Sensory Studies*, vol 6, p. 131-144.
- Dailliant-Spinnler B., Issanchou S. (1995). Influence of label and location of testing on acceptability of cream cheese varying in fat content. *Appetite*, vol 24, p. 101-105.
- De Castro J.M. (1990). Social facilitation of duration and size but not rate of the spontaneous meal intake of humans. *Physiology and Behavior*, vol 1129-1135.
- De Graaf C., Cardello A.V., Kramer F.M., Leshner L.L., Meiselman H.L., Schutz H.G. (2005). A comparison between liking ratings obtained under laboratory and field conditions: the role of choice. *Appetite*, vol 44, p. 15-22.
- Deci E.L., Eghrari H., Patrick B.C., Leone D. (1994). Facilitating internalization: the self-determination theory perspective. *Journal of Personality*, vol 62, p. 119-142.
- Delarue J., Rason J., Venturin C., Sieffermann J.M. (2001). Influence of the product presentation procedure on the outcome of hedonic tests: Is simultaneous presentation an interesting

- alternative to monadic sequential? In. *4th Pangborn sensory science symposium: A sense odyssey*, Dijon, France.
- Deldin P.J., Levin I.P. (1986). The effect of mood induction in a risky decision-making task. *Bulletin of the Psychonomic Society*, vol 24, p. 4-6.
- Deliza R., MacFie H. (1996). The generation of sensory expectations by external cues and its effects on sensory perception and hedonic ratings: a review. *Journal of Sensory Studies*, vol 11, p. 103-128.
- Dember W., Earl R. (1957). Analysis of exploratory, manipulatory and curiosity behaviours. *Psychological Review*, vol 64, p. 91-96.
- Dickson P.R., Sawyer A.G. (1990). The price knowledge and search of supermarket shoppers. *Journal of Marketing*, vol 54, p. 42-53.
- Dransfield E., Zamora F., Bayle M.C. (1998). Consumer selection of steaks as influenced by information and price index. *Food Quality and Preference*, vol 9, p. 321-326.
- Drewnowski A., Grinker J.A., Hirsch J. (1987). Obesity and flavor perception: multidimensional scaling of soft drinks. *Appetite*, vol 3, p. 361-368.
- Drifford V., Lemmet J., Bernard P., Laporte V., Chambres P. (1995). Acceptance tests and memory of the context of consumption: analysis and experimentation. *Viandes et Produits Carnes*, vol 16, p. 107-111.
- Eagly A.H., Chaiken S. (1993). *The psychology of attitudes*. Orlando : Hartcourt Brace & Company.
- Earthy P.J., MacFie H., Hedderley D.I. (1997). Effect of question order on sensory perception and preference in central location trials. *Journal of Sensory Studies*, vol 12, p. 215-237.
- Edwards J., Meiselman H.L., Edwards A., Leshner L.L. (2003). The influence of eating location on the acceptability of identically prepared foods. *Food Quality and Preference*, vol 14, p. 647-652.
- Eindhoven J., Peryam D.R. (1959). Measurement of preferences for food combinations. *Food Technology*, vol 13, p. 379-382.
- Engell D., Kramer F.M., Malafi T., Salomon M., Leshner L.L. (1996). Effects of effort and social modeling on drinking in humans. *Appetite*, vol 26, p. 129-138.
- Engell D., Bordi P., Borja M., Lambert C., Rolls B.J. (1998). Effects of information about fat content on food preferences in pre-adolescent children. *Appetite*, vol 30, p. 269-282.
- Esposito J., Jobe J. (1991). A general model of the survey interaction process. In. *Seventh Annual Research Conference of the Bureau of the Census*, Washington, DC, United-States.
- Fantino M. (1992). Etat nutritionnel et perception affective de l'aliment. In. Giachetti I. (Ed.) *Plaisir et préférences alimentaires* (p. 33-45). Paris : Polytechnica.

- Feldman J.M., Lynch J.G. (1988). Self-generated validity and other effects of measurements on belief, attitude, intention and behavior. *Journal of Applied Psychology*, vol 73, p. 421-435.
- Feunekes G.I., De Graaf C., Van Staveren W.A. (1995). Social facilitation of food intake is mediated by meal duration. *Physiology and Behavior*, vol 58, p. 551-558.
- Filipello F. (1956). Factors in the analysis of man panel wine-preferences data. *Food Technology*, vol 10, p. 321-326.
- Finch J. (1987). The vignette technique in survey research. *Sociology*, vol 21, p. 105-114.
- Fishbein M., Ajzen I. (1975). *Belief, attitude, intention and behavior: an introduction to theory and research*. Reading, Mass : Addison-Wesley.
- Folkes V.S. (1984). Consumer reactions to product failure: an attributional approach. *Journal of Consumer Research*, vol 14, p. 398-409.
- Forde C.G., Delahunty C.M. (2004). Understanding the role cross-modal sensory interactions play in food acceptability in younger and older consumers. *Food Quality and Preference*, vol 15, p. 715-727.
- Fox J.A. (1995). Determinants of consumer acceptability of bovine somatotropin. *Review of Agricultural Economics*, vol 17, p. 51-62.
- Fox J.A., Shogren J.F., Hayes D.J., Kliebenstein J.B. (1998). CVM-X: calibrating contingent values with experimental auction markets. *American Journal of Agricultural Economics*, vol 80, p. 455-465.
- Frandsen L.W., Dijksterhuis G.B., Brockhoff P., Nielsen J.H., Martens M. (2003). Subtle differences in milk: comparison of an analytical and an affective test. *Food Quality and Preference*, vol 14, p. 515-526.
- Frandsen L.W., Dijksterhuis G.B., Brockhoff P., Nielsen J.H., Martens M. (2007). Feeling as a basis for discrimination: comparison of a modified authenticity test with the same-different test for slightly different types of milk. *Food Quality and Preference*, vol 18, p. 97-105.
- Furst T., Connors M., Bisogni C.A., Sobal J., Falk L. (1996). Food choice: a conceptual model of the process. *Appetite*, vol 26, p. 247-265.
- Gains N. (1989). The integration of personal construct theory in food acceptability research. Thesis, University of Reading, UK.
- Garber L.L., Hyatt E.M., Starr R.G. (2003). Measuring consumer response to food products. *Food Quality and Preference*, vol 14, p. 3-15.
- Gordon N.M., Norback J.P. (1985). Choosing objective measures when using sensory methods for optimization and product positioning. *Food Technology*, vol 39, p. 96-101.

- Green D.M., Butts J.S. (1945). Factors affecting acceptability of meals served in the air. *Journal of the American Dietetic Association*, vol 21, p. 415-419.
- Green P.E. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research*, vol 5, p. 103-123.
- Griffin R., Stauffer L. (1990). Product optimization in central-location testing and subsequent validation and calibration in home-use testing. *Journal of Sensory Studies*, vol 5, p. 231-240.
- Gross E.J. (1964). The effect of question sequence on measures of buying interest. *Journal of Advertising Research*, vol 4, p. 40-41.
- Grunert K.G. (2003). Purchase and consumption: the interdisciplinary nature of analysing food choice. *Food Quality and Preference*, vol 14, p. 39-40.
- Guevara-Aguilar R., Jimenez-Montufar L.L., Garcia-Diaz D.E. (1988). Olfactory and visceral projections to the paraventricular nucleus. *Brain Research Bulletin*, vol 20, p. 799-801.
- Hayes D.J., Shogren J.F., Shin S.U., Kliebenstein J.B. (1995). Valuing food safety in experimental auction markets. *American Journal of Agricultural Economics*, vol 77, p. 40-53.
- Hedderley D.I., Meiselman H.L. (1995). Modelling meal acceptability in a free choice environment. *Food Quality and Preference*, vol 6, p. 15-26.
- Helgesen H., Solheim R., Naes T. (1998). Consumer purchase probability of dry fermented lamb sausages. *Food Quality and Preference*, vol 9, p. 295-301.
- Helleman U., Tuorila H.M. (1991). Pleasantness ratings and consumption of open sandwiches with varying NaCl and acid contents. *Appetite*, vol 17, p. 229-238.
- Helleman U., Mela D.J., Aaron J.I., Evans R.E. (1992). Role of fat in meal acceptance. In. *Pangborn Symposium*, Helsinki, Finland.
- Hersleth M., Mevik B.H., Naes T., Guinard J.X. (2003). Effect of contextual factors on liking for wine - use of robust design methodology. *Food Quality and Preference*, vol 14, p. 615-622.
- Hersleth M., Ueland O., Allain H., Naes T. (2005). Consumer acceptance of cheese, influence of different testing conditions. *Food Quality and Preference*, vol 16, p. 103-110.
- Hetherington M.M., Bell A., Rolls B.J. (2000). Effects of repeat consumption on pleasantness, preference and intake. *British Food Journal*, vol 102, p. 507-521.
- Hicks J.R. (1945). The generalized theory of consumer surplus. *Review of Economic Studies*, vol 13, p. 68-73.
- Hill S.W. (1974). Eating responses of humans during dinner meals. *Journal of Comparative and Physiological Psychology*, vol 86, p. 652-657.

- Hoch S.J. (1984). Availability and interference in predictive judgment. *Journal of Experimental Psychology Learning Memory and Cognition*, vol 10, p. 649-662.
- Hoffman E., Menkhaus D.J., Chakravarti D., Field R.A., Whipple G.D. (1993). Using laboratory experimental auctions in marketing research: a case study of new packaging for fresh beef. *Marketing Science*, vol 12, p. 318-338.
- Horowitz J.K., McConnell K.E. (2002). A review of WTA/WTP studies. *Journal of Environmental Economics and Management*, vol 44, p. 426-447.
- Hoyer S., Simchem U.,Zunft H.J.F. (2003). A vending machine to measure choice and consumption by different age groups in real-life situation. In. *The 5th Pangborn sensory symposium: A sensory revolution*, Boston, United States.
- Isen A.M., Shalke T.E., Clark M., Karp L. (1978). Affect, accessibility of material in memory, and behavior: a cognitive loop? *Journal of Personality and Social psychology*, vol 36, p. 1-12.
- Issanchou S., Hossenlopp P. (1992). Les mesures hédoniques : méthodes, portées et limites. In. Giachetti I. (Ed.) *Plaisirs et préférences alimentaires*, Paris : Polytechnica.
- Izard C.E., Dougherty F.E., Bloxom B.M., Kotsch W.E. (1974). *The differential emotions scale: A method of measuring the subjective experience of discrete emotions*. Department of Psychology. Nashville, TN: Vanderbilt University.
- Jaeger S.R. (2000). Uncovering cultural differences in choice behaviour between Samoan and New Zealand consumers: a case study with apples. *Food Quality and Preference*, vol 11, p. 405-417.
- Jaeger S.R., Andani Z., Wakeling I.N., MacFie, H.J.H. (1998). Consumer preferences for fresh and aged apples: a cross-cultural comparison. *Food Quality and Preference*, vol 9, p. 355-366.
- Jaeger S.R., Meiselman H.L. (2004). Perceptions of meal convenience: the case of at-home evening meals. *Appetite*, vol 42, p. 317-325.
- Jaeger S.R., Lusk J.L., House L.O., Valli C., Moore M., Morrow B., Traill W.B. (2004). The use of non-hypothetical experimental markets for measuring the acceptance of genetically modified foods. *Food Quality and Preference*, vol 15, p. 701-714.
- Jones L.V., Peryam D.R., Thurstone L.L. (1955). Development of a scale for measuring soldiers' food preferences. *Food Research*, vol 20, p. 512-520.
- Kähkönen P., Tuorila H.M., Lawless H. (1997). Lack of effect of taste and nutrition claims on sensory and hedonic responses to a fat-free yogurt. *Food Quality and Preference*, vol 8, p. 125-130.
- Kalwani M.U., Silk A.J. (1982). On the reliability and predictive validity of purchase intention measures. *Marketing Science*, vol 1, p. 243-286.
- Kamen J., Peryam D.R., Peryam D.R., Kroll B.J. (1969). Hedonic differences as a function of number of samples evaluated. *Journal of Food Science*, vol 34, p. 475-479.

- Kamenetzky J., Pilgrim F.J., Schutz H.G. (1957). Relationship of consumption to preference under different field conditions. Quatermaster Food and Container Interim Report.
- Khan M.A. (1981). Evaluation of food selection patterns and preferences. *CRC Critical Reviews in Food Science and Nutrition*, vol 15, p. 129-153.
- Kanarek R.B., Orthen-Gambill N. (1986). Complex interactions affecting nutrition-behavior research. *Nutrition Reviews*, vol 44, p. 172-175.
- Kihlstrom J.F. (1987). The cognitive unconscious. *Science*, vol 237, p. 1445-1452.
- King S.C., Weber A.J., Meiselman H.L., Lv N. (2004). The effect of meal situation, social interaction, physical environment and choice on food acceptability. *Food Quality and Preference*, vol 15, p. 645-653.
- King S.C., Meiselman H.L., Hottenstein A.W., Work T.M., Cronk V. (2007). The effects of contextual variables on food acceptability: a confirmatory study. *Food Quality and Preference*, vol 18, p. 58-65.
- Knetsch J., Tang F.F., Thaler, R. (1998). The endowment effect and repeated market trials: is the Vickrey auction demand revealing? Simon Fraser University. Unpublished Work.
- Kotler P. (1974). Marketing Management: analysis, planning, implementation and control. *Revue Française du Marketing*, vol 54, p. 13-29.
- Kozłowska K., Jeruszka M., Matuszewska I., Roszkowski W., Pikielna N., Brzozowska A. (2003). Hedonic tests in different locations as predictors of apple juice consumption at home in elderly and young subjects. *Food Quality and Preference*, vol 14, p. 653-666.
- Köster E.P. (1981a). Sensory evaluation in a natural environment. In. Schreier P. (Ed.) *Flavour'81* (p. 93-100). Berlin : Walter de Gruyter and Co.
- Köster E.P. (1981b). Time and frequency analysis: a new approach to the measurement of some less-well-known aspects of food preferences. In. Solms J., Hall R.L. (Eds.) *Criteria of food acceptance* (p. 240-252). Zurich : Forster Verlag AG.
- Köster E.P. (1996). The consumer? The quality? In. *Production industrielle et qualité sensorielle; Agoral 96. Huitième rencontres scientifiques et technologiques des industries alimentaires*, Dijon, France.
- Köster E.P. (1998). Les épreuves hédoniques. In. SSHA. (Ed.) *Evaluation sensorielle, manuel méthodologique* (p. 182-203). Paris : Lavoisier, Tec et Doc.
- Köster E.P. (2003). The psychology of food choice: some often encountered fallacies. *Food Quality and Preference*, vol 14, p. 359-373.
- Köster, M.A., Prescott J., Köster E.P. (2004). Incidental learning and memory for three basic tastes in food. *Chemical Senses*, vol 29, p. 441-453.

- Köster E.P., Mojet J. (2006). Boredom and the reasons why new products fail. In: MacFie H. (Ed.) *Consumer led food product development*. Abbingdon Cambridge : Woodhead Publishing.
- Kramer F.M., Rock K., Engell D. (1992). Effect of time of day and appropriateness on food intake and hedonic ratings at morning and midday. *Appetite*, vol 18, p. 1-13.
- Kramer F.M., Leshner L.L., Meiselman H.L. (2001). Monotony and choice: repeated serving of the same item to soldiers under field conditions. *Appetite*, vol 36, p. 239-240.
- Krosnick J. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, vol 5, p. 213-236.
- Laeng B., Berridge-Kent C., Butter C.M. (1993). Pleasantness of a sweet taste during hunger and satiety: effects of gender and "sweet tooth". *Appetite*, vol 21, p. 247-254.
- Lambert J.L. (1999). Les mangeurs face aux nouvelles technologies alimentaires. In: *L'agro-alimentaire : histoire et modernité*, Nantes, France.
- Lange C., Rousseau F., Issanchou S. (1999). Expectation, liking and purchase behaviour under economical constraint. *Food Quality and Preference*, vol 10, p. 31-39.
- Lange C., Issanchou S., Combris P. (2000). Expected versus experienced quality: trade-off with price. *Food Quality and Preference*, vol 11, p. 289-297.
- Lange C., Martin C., Chabanet C., Combris P., Issanchou S. (2002). Impact of the information provided to consumers on their willingness to pay for champagne: comparison with hedonic scores. *Food Quality and Preference*, vol 13, p. 597-608.
- Lawless H. (1983). Contextual effects in category ratings. *Journal of Testing and Evaluation*, vol 11, p. 346-349.
- Lawless H., Heymann H. (1999). *Sensory evaluation of food. Principles and practices*. Gaithersburg, Maryland : Aspen Publishers, Inc.
- Lähteenmäki L., Tuorila H.M. (1994). Liking for ice cream measured with three procedures: side-by-side, after consumption and single samples. *Journal of Sensory Studies*, vol 9, p. 455-465.
- Lähteenmäki L., Tuorila H.M. (1995). Consistency of liking and appropriateness ratings and their relation to consumption in a product test of ice cream. *Appetite*, vol 25, p. 189-198.
- Lähteenmäki L., Tuorila H.M. (1997). Item-by-use appropriateness of drinks varying in sweetener and fat content. *Food Quality and Preference*, vol 8, p. 85-90.
- Lähteenmäki L., Tuorila H.M. (1998). Predicting the intention to use juice or milk in three contexts. *Food Quality and Preference*, vol 9, p. 231-236.
- Lemoine J.F. (1999). L'influence de la situation d'utilisation du produit sur la recherche d'information externe du consommateur. *Direction et Gestion, La Revue des Sciences de Gestion*, vol 177, p. 39-51.

- Leon F., Couronne T., Marcuz M.C., Köster E.P. (1999). Measuring food liking in children: a comparison of non verbal methods. *Food Quality and Preference*, vol 10, p. 93-100.
- Levitz L.S. (1975). The susceptibility of human feeding behavior to external control. *Obesity in perspective*. Washington DC : Government Printing Office.
- Levy C.M. (1998). La dynamique des préférences: application aux boissons.Thèse, Université de Bourgogne, Dijon, France.
- Levy C.M., Köster E.P. (1999). The relevance of initial hedonic judgements in the prediction of subtle food choices. *Food Quality and Preference*, vol 10, p. 185-200.
- Lévy C.M., MacRae A.W., Köster E.P. (2006). Perceived stimulus complexity and food preference. *Acta Psychologica*, vol 123, p. 394-413.
- List J.A., Shogren J.F. (1998). Calibration of the differences between actual and hypothetical valuations in a field experiment. *Journal of Economic Behavior and Organization*, vol 37, p. 193-205.
- Lucas F., Bellisle F. (1987). The measurement of food preferences in humans: do taste and spit tests predict consumption? *Physiology and Behavior*, vol 39, p. 739-743.
- Lusk J.L. (2003). Effect of cheap talk on consumer willingness to pay for Golden Rice. *American Journal of Agricultural Economics*, vol 85, p. 840-856.
- Lusk J.L., Daniel M.S., Lusk C.L., Mark D.R. (2001). Alternative calibration and auction institutions for predicting consumer willingness to pay for non-genetically modified corn chips. *Journal of Agricultural and Resource Economics*, vol 26, p. 40-57.
- Lusk J.L., Fox J.A., Schroeder T.C., Mintert J., Koohmaraie M. (2001). In-store valuation of steak tenderness. *American Journal of Agricultural Economics*, vol 83, p. 539-550.
- Lusk J.L., Fox J.A. (2003). Value elicitation in retail and laboratory environments. *Economics Letters*, vol 79, p. 27-34.
- Lusk J.L., Roosen J., Fox J.A. (2003). Demand for beef from cattle administered growth hormones or fed genetically modified corn: a comparison of consumers in France, Germany, the United Kingdom, and the United States. *American Journal of Agricultural Economics*, vol 85, p. 16-29.
- Lusk J.L., Schroeder T.C. (2004). Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *American Journal of Agricultural Economics*, vol 86, p. 467-482.
- Lusk J.L., Feldkamp T., Schroeder T.C. (2004). Experimental auction procedure: impact on valuation of quality differentiated goods. *American Journal of Agricultural Economics*, vol 86, p. 389-405.
- Maller O., Dubose C.N., Cardello A.V. (1980). Consumer opinions of hospital food and foodservice. *Journal of the American Dietetic Association*, vol 76, p. 236-242.

- Marette, S., Roosen, J., Blanchemanche, S., and Verger, P. (2006). Health information and the choice of fish species: an experiment measuring the impact of risk and benefit information. Working Paper 06-WP 421. Unpublished Work
- Mathieu J.P. (2002). La représentation d'un produit: une histoire de contexte. In. *Congrès de l'Association Française de Marketing*, Lille, France.
- Matuszewska I., Barylko-Pikielna N., Szczecinska A., Radzanovska J. (1997). Comparison of three procedures for consumer assessment of fat spreads : Short report. *Polish Journal of Food and Nutrition Science*, vol 6, p. 139-142.
- McBride R.L. (1982). Range bias in sensory evaluation. *Journal of Food Technology*, vol 17, p. 405-410.
- McBride R.L. (1985). Stimulus range influences intensity and hedonic ratings of flavor. *Appetite*, vol 6, p. 125-131.
- McBride R.L. (1986). Hedonic rating of food: single or side-by-side sample presentation? *Journal of Food Technology*, vol 21, p. 355-363.
- McBride R.L.M., MacFie H.J.H. (1990). *Psychological basis of sensory evaluation*. London: Elsevier Applied Science.
- McClendon M.J., O'Brien D.J. (1988). Question order effects on the determinants of subjective well-being. *Public Opinion Quarterly*, vol 52, p. 351-364.
- McDaniel M.R., Sawyer F.M. (1981). Preference testing of whiskey sour formulations: magnitude estimation versus the 9-point hedonic. *Journal of Food Science*, vol 46, p. 182-185.
- McEwan J.A. (1997). A comparative study of three product acceptability trials. *Food Quality and Preference*, vol 8, p. 183-190.
- McEwan J.A., Thomson D.M.H. (1988). A behavioural interpretation of food acceptability. *Food Quality and Preference*, vol 1, p. 3-9.
- Meilgaard M., Civille G.V., Carr B.T. (1991). *Sensory evaluation techniques*. Boca Raton, Florida : CRC Press, LLC.
- Meiselman H.L. (1992). Methodology and theory in human eating research. *Appetite*, vol 19, p. 49-55.
- Meiselman H.L. (1996). The contextual basis for food acceptance, food choice and food intake. In. Meiselman H.L., MacFie H. (Eds.) *Food choice acceptance and consumption* (p. 139-263). Glasgow : Blackie Academic and Professional.
- Meiselman H.L. (2003). A three-factor approach to understanding food quality: the product, the person and the environment. *Food Service Technology*, vol 3, p. 99-105.

- Meiselman H.L., Hirsch E.S., Popper R.D. (1988). Sensory, hedonic and situational factors in food acceptance consumption. In. Thomson D.M.H. (Ed.) *Food acceptability* (p. 77-87). London : Elsevier.
- Meiselman H.L., Hedderley D.I., Staddon S.L., Pierson B.J., Symonds C.R. (1994). Effect of effort on meal selection and meal acceptability in a student cafeteria. *Appetite*, vol 23, p. 43-55.
- Meiselman H.L., De Graaf C., Leshner L.L. (2000). The effects of variety and monotony on food acceptance and intake at a midday meal. *Physiology and Behavior*, vol 70, p. 119-125.
- Meiselman H.L., Johnson J.L., Reeve W.G., Crouch J.E. (2000). Demonstrations of the influence of the eating environment on food acceptance. *Appetite*, vol 35, p. 231-237.
- Meiselman H.L., Schutz H.G. (2003). History of food acceptance research in the US Army. *Appetite*, vol 40, p. 199-216.
- Meiselman H.L., King S.C., Herniques A. (2006). The challenge of using psychographics to segment consumer panels. In. *A Sense of Diversity - 2nd European Conference on Sensory Consumer Science of Food and Beverages*. The Hague, NL.
- Mela D.J. (1989). A comparison of single and concurrent evaluations of sensory and hedonic attributes. *Journal of Food Science*, vol 54, p. 1098-1100.
- Mela D.J., Rogers P.J., Shepherd R., MacFie H. (1992). Commentary. real people, real foods, real eating situations: real problems and real advantages. *Appetite*, vol 19, p. 69-73.
- Melton B.E., Huffman W.E., Shogren J.F., Fox J.A. (1996). Consumer preferences for fresh food items with multiple quality attributes: evidence from an experimental auction of pork chops. *American Journal of Agricultural Economics*, vol 78, p. 916-923.
- Meyers A.W., Stunkard A.J., Coll M. (1980). Food accessibility and food choice. *Archives of General Psychiatry*, vol 37, p. 1133-1135.
- Miller K.E., Ginter J.L. (1979). An investigation of situational variation in brand choice behavior and attitude. *Journal of Marketing Research*, vol 16, p. 111-123.
- Miller P.G., Nair J.H., Harriman A.J. (1955). A household and a laboratory type of panel for testing consumer preference. *Food Technology*, vol 9, p. 445-449.
- Mojet J., Köster E.P. (1986). Research on the appreciation of three low-alcohol beers. Report Psychological Laboratory, Utrecht University, Unpublished.
- Mojet J., Köster E.P. (2005). Sensory memory and food texture. *Food Quality and Preference*, vol 16, p. 251-266.
- Monneuse M.O., Bellisle F., Louis-Sylvestre J. (1991a). Responses to an intense sweetener in humans: immediate preference and delayed effects on intake. *Physiology and Behavior*, vol 49, p. 325-330.

- Monneuse M.O., Bellisle F., Louis-Sylvestre J. (1991b). Impact of sex and age on sensory evaluation of sugar and fat in dairy products. *Physiology and Behavior*, vol 50, p. 1111-1117.
- Moorthy K.S. (1991). Measuring overall judgments and attribute evaluations: overall first vs. attributes first. Technical Working Paper: Marketing Science Institute.
- Moskowitz H.R. (1995). The dollar value of product quality: the effect of pricing versus overall liking on consumer stated purchase intent for pizza. *Journal of Sensory Studies*, vol 10, p. 239-247.
- Moskowitz H.R. (1997). Base size in product testing: a psychophysical viewpoint and analysis. *Food Quality and Preference*, vol 8, p. 247-255.
- Moskowitz H.R. (2000). Engineering out food boredom: a product development approach that combines home use tests and time-preference analysis. *Food Quality and Preference*, vol 11, p. 445-456.
- Moskowitz H.R., Muñoz A.M., Gacula M.C. (2003). *Viewpoints and controversies in sensory science and consumer product testing*. Trumbull : Food and Nutrition Press, Inc.
- Møller P., Wulff C., Köster E.P. (2004). Do age differences in odour memory depend on differences in verbal memory? *Neuroreport*, vol 15, p. 915-917.
- Muller K. (2004). Prédiction de l'évolution des préférences des consommateurs dans le temps en fonction de facteurs sensoriels et plus particulièrement de la complexité perçue et de la familiarité et de facteurs individuels: application à deux gammes de produits. Thèse, INAPG, Paris, France.
- Murphy E.F., Clark B.S., Berglund R.M. (1958). A consumer survey versus panel testing for acceptance evaluation of Main sardines. *Food Technology*, vol 12, p. 222-226.
- Nantachai K., Petty M.F., Scriven F.M. (1991). An application of contextual evaluation to allow simultaneous food product development for domestic and export markets. *Food Quality and Preference*, vol 3, p. 13-22.
- Nossanchuck T.A. (1972). The vignette as an experimental approach to the study of social status; an exploratory study. *Social Science Research*, vol 1, p. 107-120.
- Noussair C., Robin S., Ruffieux B. (2004a). A comparison of hedonic rating and demand-revealing auctions. *Food Quality and Preference*, vol 15, p. 393-402.
- Noussair C., Robin S., Ruffieux B. (2004b). Revealing consumers' willingness-to-pay: A comparison of the BDM mechanism and the Vickrey auction. *Journal of Economic Psychology*, vol 25, p. 725-741.
- Nunes J., Boatwright P. (2004). Incidental prices and their effect on willingness to pay. *Journal of Marketing Research*, vol 16, p. 457-466.

- Olson J.C. (1981). The importance of cognitive processes and existing knowledge structures for understanding food acceptance. In. Solms J., Hall R.L. (Eds.) *Criteria of Food Acceptance* (p. 69-81). Zurich : Foster-Verlag.
- Pavlov I.P. (1927). *Conditioned reflexes. An investigation of the physiological activity of the cerebral cortex*. Oxford : Oxford University Press.
- Perez C., Dalix A.M., Guy-Grand B., Bellisle F. (1994). Human responses to five concentrations of sucrose in a dairy product: immediate and delayed palatability effects. *Appetite*, vol 23, p. 165-178.
- Perrien J., Cheron E.J., Zins M. (1984). Méthodes d'échantillonnage. *Recherche en marketing: méthodes et décisions* (p. 203-241). Montreal : Gâetan Morin.
- Peryam D.R., Pilgrim F.J. (1957). Hedonic scale method of measuring food preferences. In. *Food technology symposium*, p. 9-14.
- Peryam D.R., Polemis B.W., Kamen J., Eindhoven J., Pilgrim F.J. (1960). *Food preferences of men in the armed forces*. Chicago, Quatermaster Food and Container Institute.
- Petit C., Sieffermann J.M. (2007). Testing consumer preferences for iced-coffee: Does the drinking environment have any influence? *Food Quality and Preference*, vol 18, p. 161-172.
- Pieron H. (1974). *La sensation* . Paris : Presse Universitaire de France.
- Pilgrim F.J. (1957). The components of food acceptance and their measurement. *American Journal of Clinical Nutrition*, vol 5, p. 171-175.
- Pilgrim F.J., Kamen J. (1963). Predictors of human food consumption. *Science*, vol 139, p. 501-502.
- Pliner P. (1982). The effect of mere exposure on liking for edible substances. *Appetite*, vol 2, p. 283-290.
- Pliner P. (1992). Let's not throw out the barley with the dishwater: comments on Meiselman's: "Methodology and theory in human eating research". *Appetite*, vol 19, p. 74-75.
- Pliner P., Bell R., Kinchla M., Hirsch E.S. (2003). Time to eat: the impact of time and social facilitation on food intake.
- Popper R.D., Maller O., Cardello A.V. (1989). Hedonic ratings and consumption as measures of preference for a salted entree. *Chemical Senses*, vol 14, p. 739.
- Popper R.D., Rosenstock W., Schraidt M., Kroll B.J. (2004). The effect of attribute questions on overall liking ratings. *Food Quality and Preference*, vol 15, p. 853-858.
- Porcherot C., Issanchou S. (1998). Dynamics of liking for flavoured crackers: test of predictive value of a boredom test. *Food Quality and Preference*, vol 9, p. 21-29.

- Posri W., MacFie H., Henson S. (2001). Improving the predictability of consumer liking from central location test in tea. In. *Pangborn Memorial Symposium: a sense odyssey*, Dijon, France.
- Pound C., Duizer L., McDowell K. (2000). Improved consumer product development. I. Is a laboratory necessary to assess consumer opinion? *British Food Journal*, vol 102, p. 810-820.
- Raats M.M., Shepherd R. (1991). An evaluation of the use and perceived appropriateness of milk using the repertory grid method and the "item-by-use" appropriateness method. *Food Quality and Preference*, vol 3, p. 89-100.
- Randall E., Sanjur D. (1981). Food preferences - Their conceptualization and relationship to consumption. *Ecology of Food and Nutrition*, vol 11, p. 151-161.
- Resurreccion A.V. (1998). *Consumer sensory testing for product development*. Gaithersburg, Maryland : Aspen Publishers Inc.
- Riskey D.R., Parducci A., Beauchamp G.K. (1979). Effects of context in judgments of sweetness and pleasantness. *Perception and Psychophysics*, vol 26, p. 171-176.
- Rolls B.J. (1987). Sweetness and satiety. In. Dobbing (Ed.) *Sweetness* (p. 161-173). London : Springer-Verlag.
- Rolls B.J., Duijvenvoorde P.M.V., Rolls E.T. (1984). Pleasantness changes and food intake in a varied four-course meal. *Appetite*, vol 348.
- Rolls B.J., Hetherington M.M. (1990). A behavioral scientist's perspectives on the study of diet and behavior. In. Anderson G.H. (Ed.) *Diet and Behavior* (p. 209-219). London : Springer Verlag.
- Rolls B.J., Shide D.J. (1992). Both naturalistic and laboratory-based studies contribute to the understanding of human eating behavior. *Appetite*, vol 19, p. 76-77.
- Roth D.A., Herman C.P., Polivy J., Pliner P. (2001). Self-presentational conflict in social eating situations: a normative perspective. *Appetite*, vol 36, p. 165-171.
- Rozin P. (1990). Acquisition of stable food preferences. *Nutrition Reviews*, vol 48, p. 106-113.
- Rozin P. (1996). The socio-cultural context of eating and food choice. In. Meiselman H.L., MacFie H. (Eds.) *Food choice acceptance and consumption* (p. 83-104). Glasgow : Blackie Academic and Professional.
- Rozin P., Markwith M. (1991). Cross-domain variety seeking in human food choice. *Appetite*, vol 16, p. 57-59.
- Rozin P., Tuorila H.M. (1993). Simultaneous and temporal contextual influences on food acceptance. *Food Quality and Preference*, vol 4, p. 11-20.
- Rust R.T., Moorman C., Dickson P.R. (2002). Getting return on quality: revenue expansion, cost reduction or both? *Journal of Marketing*, vol 66, p. 7-24.

- Ryan R.M. (1982). Control and information in the intrapersonal sphere: an extension of cognitive evaluation theory. *Journal of Personality and Social psychology*, vol 43, p. 450-461.
- Ryynanen S., Tuorila H.M., Hyvonen L. (2001). Perceived temperature effects on microwave heated meals and meal components. *Food Service Technology*, vol 1, p. 141-148.
- Sanogo D., Masters W.A. (2002). A market-based approach to child nutrition: mothers' demand for quality certification of infant foods in Bamako, Mali. *Food Policy*, vol 27, p. 251-268.
- Sauerhoff K., Gualtieri T., Brumbaugh K., Craig-Petsinger D. (2005). The application of equivalence testing to consumer research where the objective is parity. In. *6th Pangborn Sensory Science Symposium*, Harrogate, UK.
- Schaefer E.E. (1979). Manual on consumer sensory evaluation. In. *American Society for Testing and Materials*, Philadelphia: ASTM.
- Schuman J., Presser S. (1981). *Questions and answers in attitude surveys: experiments in question form, wording and context*. New York : Academic Press.
- Schutz H.G. (1957). Preference ratings as predictors of food consumption, *American Psychologist*, volume 12, p. 380.
- Schutz H.G. (1965). A food action rating scale for measuring food acceptance. *Journal of Food Science*, vol 30, p. 365-374.
- Schutz H.G. (1988). Beyond preference: appropriateness as a measure of contextual acceptance of food. In. Thomson D.M.H. (Ed.) *Food Acceptability* (p. 115-134). New York : Elsevier.
- Schutz H.G., Ortega J.H. (1974). Consumer attitudes toward wine. *American Journal of Enology and Viticulture*, vol 25, p. 33-38.
- Schutz H.G., Rucker M.H., Russel G.F. (1975a). Food and food use classification. *Food Technology*, vol 29, p. 50-64.
- Schutz H.G., Fridgen J.D., Damrell J.D. (1975b). Consumer perceptions of rice and related products. *Journal of Food Science*, vol 40, p. 277-281.
- Schutz H.G., Moore S.M., Rucker M.H. (1977). Predicting food purchase and use by multivariate attitudinal analysis. *Food Technology*, vol 29, p. 50-64.
- Schwartz N., Pratt C.H. (1956). Simultaneous vs. successive presentation in a paired comparison situation. *Food Research*, vol 21, p. 103-108.
- Scriven F.M., Gains N., Green S.R., Thomson D.M.H. (1989). A contextual evaluation of alcoholic beverages using the repertory grid method. *International Journal of Food Science and Technology*, vol 24, p. 173-182.
- Sharp W.F., Norback J.P., Stuibler D.A. (1986). Using a new measure to define shelf life of fresh whitefish. *Journal of Food Science*, vol 51, p. 936-939.

- Shepherd R. (1985). Factors influencing food choice. *Nutrition and Food Science*, vol 96, p. 10-11.
- Shepherd R., Griffiths N.M. (1987). Preferences for eggs produced under different systems assessed by consumer and laboratory panels. *Lebensmittel Wissenschaft und Technologie*, vol 20, p. 128-132.
- Shepherd R., Farleigh C.A., Wharf S.G. (1991). Effect of quantity consumed on measures of liking for salt concentrations in soup. *Journal of Sensory Studies*, vol 4, p. 238.
- Shiv B., Fedorikhin A. (1999). Heart and mind in conflict: the interplay of affect and cognition in consumer decision making. *Journal of Consumer Research*, vol 26, p. 278-292.
- Shogren J.F., Margolis M., Koo C., List J.A. (2001). A random nth-price auction. *Journal of Economic Behavior and Organization*, vol 46, p. 409-421.
- Sidel J.L., Stone H., Wollsey A., Macredy J.M. (1972). Correlation between hedonic ratings and consumption of beer. *Journal of Food Science*, vol 37, p. 335-335.
- Siegel S.F., Risvik E. (1987). Cognitive set and food acceptance. *Journal of Food Science*, vol 52, p. 825-826.
- Simon H.A. (1956). Rational choice and the structure of the environment. *Psychological Review*, vol 63, p. 129-152.
- Simone E.M., Pangborn R.M. (1957). Consumer acceptance methodology: one vs. two samples. *Food Technology*, vol 11, p. 25-29.
- Sosa M., Martinez C., Arruiz F., Hough G., Mucci A. (2005). Degree of appropriateness and frequency of consumption of mayonnaise, ketchup, mustard and similar sauces in Argentina. *Food Quality and Preference*, vol 16, p. 667-674.
- Stang D. (1975). Effect of "mere exposure" on learning and affect. *Journal of Personality and Social psychology*, vol 31, p. 7-12.
- Steenkamp J.B. (1993). Food consumption behavior. *European Advances in Consumer Research*, vol 1, p. 401-409.
- Stefani G., Romano D., Cavicchi A. (2006). Consumer expectations, liking and willingness to pay for specialty foods: do sensory characteristics tell the whole story? *Food Quality and Preference*, vol 17, p. 53-62.
- Stolte J.F. (1994). The context of satisficing in vignette research. *The Journal of Social Psychology*, vol 134, p. 727-733.
- Stone H., Sidel J.L. (1993). *Sensory evaluation practices, 2nd edition*. San Diego, California : Academic Press Inc.

- Strack F. (1992). Order effects in survey research: activation and information functions of preceding questions. In. Schwarz, N., Sudman, S (Eds.) *Context effects in social and psychological research* (p. 23-34), New York : Springer-Verlag.
- Stroebele N., De Castro J.M. (2004). Effect of ambience on food intake and food choice. *Nutrition*, vol 20, p. 821-838.
- Sulmont C. (2000). Impact de la mémoire des odeurs sur la réponse hédonique au cours d'une exposition répétée. Thèse, Université de Bourgogne, Dijon, France.
- Surprenant C.F., Solomon M.R. (1987). Predictability and personalization in the service encounter. *Journal of Marketing*, vol 51, p. 86-96.
- Taylor D.E., Pangborn R.M. (1990). Temporal aspects of hedonic responses. *Journal of Sensory Studies*, vol 4, p. 241-247.
- Tesser A. (1978). Self-generated attitude change. In. Berkowitz L. (Ed.) *Advances in experimental social psychology*. (p. 289-338). San Diego, CA : Academic Press.
- Thomson D.M.H., McEwan J.A. (1988). An application of the repertory grid method to investigate consumer perceptions of food. *Appetite*, vol 10, p. 181-193.
- Tourangeau R., Rasinski K.A., Bradburn N.M., Dandrade R. (1989). Belief accessibility and context effects in attitude measurement. *Journal of Experimental Social Psychology*, vol 25, p. 401-421.
- Tulving E. (1976). Rôle de la mémoire sémantique dans le stockage et la récupération de l'information épisodique. *Bulletin de Psychologie*, vol 30, p. 19-25.
- Tulving E. (1984). Precis of elements of episodic memory. *Behavioral and Brain Sciences*, vol 7, p. 223-268.
- Tuorila H.M., Lehtovaara A., Matuszewska I. (1990). Sandwiches and milk with varying fat and sodium contents: what is the best combination? *Food Quality and Preference*, vol 2, p. 223-231.
- Tuorila H.M., Lähtenmäki L. (1992). When is eating "real"? Response to Meiselman. *Appetite*, vol 19, p. 80-83.
- Tuorila H.M., Hyvoenen L., Vainio L. (1994). Pleasantness of cookies, juice and their combinations rated in brief taste tests and following ad libitum consumption. *Journal of Sensory Studies*, vol 2, p. 216.
- Tuorila H.M., Meiselman H.L., Cardello A.V., Leshner L.L. (1998). Effect of expectations and the definition of product category on the acceptance of unfamiliar foods. *Food Quality and Preference*, vol 9, p. 421-430.
- Tuorila H.M., Cardello A.V. (2002). Consumer responses to an off-flavor in juice in the presence of specific health claims. *Food Quality and Preference*, vol 13, p. 561-569.

- Turner M., Collison R. (1988). Consumer acceptance of meals and meal components. *Food Quality and Preference*, vol 1, p. 21-24.
- Tversky A., Kahneman D. (1973). Availability: a heuristic for judging frequency and probabilities. *Cognitive Psychology*, vol 5, p. 207-232.
- Tversky A., Kahneman D. (1991). Loss aversion in riskless choice: a reference-dependent model. *Quarterly Journal of Economics*, vol 106, p. 1039-1061.
- Tversky A., Slovic P., Kahneman D. (1990). The causes of preference reversal. *The American Economic Review*, vol 80, p. 204-217.
- Veinand B., Adam C., Godefroy C., Delarue J. (2006). Highlight of important product characteristics for consumers: comparison of three sensory descriptive methods. In. *A sense of Diversity - 2nd European Conference on Sensory Consumer Science of Food and Beverages*. The Hague, NL.
- Vickers Z.M., Christensen C.M., Fahrenholtz S.K., Gengler I.M. (1993). Effect of questionnaire design and the number of samples tasted on hedonic ratings. *Journal of Sensory Studies*, vol 8, p. 189-200.
- Vickers Z.M., Holton E. (1998). A comparison of taste test ratings, repeated consumption, and postconsumption ratings of different strengths of iced tea. *Journal of Sensory Studies*, vol 13, p. 199-212.
- Vickers Z.M., Mullan L., Holton E. (1999). Impact of differences in taste test ratings on the consumption of milk in both a laboratory and a foodservice setting. *Journal of Sensory Studies*, vol 262.
- Vickers Z.M., Holton E., Wang J. (2001). Effect of ideal-relative sweetness on yogurt consumption. *Food Quality and Preference*, vol 12, p. 521-526.
- Vickrey W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, vol 16, p. 8-37.
- Watson D., Clark L.A., Tellegen A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, vol 54, p. 1063-1070.
- Wertenbroch K., Skiera B. (2002). Measuring consumers' Willingness To Pay at the point of purchase. *Journal of Marketing Research*, vol 39, p. 228-241.
- West P. (1982). *Reproducing naturally occurring stories: vignettes in survey research*. Working Paper. Aberdeen: MRC Medical Sociology Unit.
- Williams A.A. (1983). Defining sensory quality in foods and beverages. *Chemistry and Industry*, vol 19, p. 740-745.
- Wilson T.D., Dunn D.S. (1986). Effects of introspection on attitude behavior consistency: analysing reasons versus focusing on feelings. *Journal of Experimental Social Psychology*, vol 22, p. 249-263.

- Wilson T.D., Schooler J.W. (1991). Thinking too much: introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social psychology*, vol 60, p. 181-192.
- Young P.T. (1948). Appetite, palatability and feeding habit: a critical review. *Psychological Bulletin*, vol 45, p. 289-320.
- Zajonc R.B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social psychology*, vol 9, p. 1-27.
- Zajonc R.B. (1980). Thinking feeling preferences need no inferences. *The American Psychologist*, vol 35, p. 151-175.
- Zajonc R.B. (1984). On primacy emotions. *The American Psychologist*, vol 39, p. 117-123.
- Zandstra E.H., De Graaf C., Trijp H.C.M., Van Staveren W.A. (1999). Laboratory hedonic ratings as predictors of consumption. *Food Quality and Preference*, vol 10, p. 411-418.
- Zandstra E.H., De Graaf C., Van Trijp H. (2000). Effects of variety and repeated in-home consumption on product acceptance. *Appetite*, vol 35, p. 113-119.
- Zandstra E.H., Weegels M.F., Van Spronsen A.A., Klerk M. (2004). Scoring or boring? Predicting boredom through repeated in-home consumption. *Food Quality and Preference*, vol 15, p. 549-557.
- Zellner D.A., Stewart W.F., Rozin P., Brown J.M. (1988). Effect of temperature and expectations on liking for beverages. *Physiology and Behavior*, vol 44, p. 61-68.