APPENDIX 1. METHODS

Details on analysis methodology

In order to compare the consensus levels between different social groups, we adapted the model proposed by some socio-psychologists. This model allows to statistically test (with an α risk of 5%) if a word was more frequently uttered by participants than if they had each picked up *w* words randomly among all uttered words, the number *w* being imposed by the researchers (Salès-Wuillemin et al. 2011). We adapted this model to a more general case and we set $\boldsymbol{w}$ as a variable instead of a constant. $\boldsymbol{w}$ varies with each respondent $\boldsymbol{i}$, taking the value of the number of words that each respondent $\boldsymbol{i}$ chose to give ($\boldsymbol{w_i}$). Words that pass the test are considered as « consensual words".

The null model we used for the test is the following:

Considering the respondent $\boldsymbol{i}$; let us call $\boldsymbol{w_i}$ the number of word he/she uttered. Let us call $\boldsymbol{W}$ the total number of different words uttered by all the respondents and $\boldsymbol{I}$ the total number of respondent in our sample. The null hypothesis (i.e. the hypothesis that must be rejected for a word to be considered "consensual") is the following: given that $\boldsymbol{W}$, $\boldsymbol{I}$ and all the $\boldsymbol{w_i}$ are known, each word is likely to be picked up by respondents with the same probability $\boldsymbol{p_i}$. This probability (random pick up with replacement) equals:

$$\frac{\binom{W-1}{wi-1}}{\binom{W}{wi}} = pi$$

Consequently, the number $\boldsymbol{N_j}$ of times that the word $\boldsymbol{j}$ is cited among $\boldsymbol{I}$ respondent can be defined by:

$$Nj = \sum_{i=1}^{I} Xi \text{ with } Xi \sim B(pi)$$

Then we test for each word $\boldsymbol{j}$ wether or not its frequency of utterance ($\boldsymbol{N_j}$) is below the $(1-\alpha)$ quantile of the null model (with a chosen risk α= 5%).

However, given the high number of test necessary (one per word), we stress the necessity to control for false positive detection rates using Holm-Bonferroni p-value adjustment technique for multiple test procedure (Holm 1979). Otherwise the proportion of false positive detection rate would be α/W.

This adjustment technique is rigorous but increases a lot the false negative detection rate (Moran, 2003). Consequently we coupled this approach with a reduction of the number of tested words (i.e. reducing $\boldsymbol{W}$) and tested only words that were uttered by at least 10% of the respondents.

Details on the categorization process

During the free listing tasks, farmers were free to cite, any word or group of words they liked with no restriction on the total number of words. This led to a great variety of uttered items. The themes were built after creating a semantical classification of words. Then emerging themes as well a theme of interest we defined to build the different categories. This step is defined by Vergés as a "merger between the researcher's own categorization system and what seems to emerge from the data" (Vergès 1992). Our categorical evaluation grid was the same across study sites and free listing tasks to allow comparing sites and representations. It was established for all items by the same researcher (CV) and crosschecked by another (RM) for

consistency. Below are some examples of aggregated categories. The weight of a category represents the number of prototypes it contains.

Table A1.1 Thematic categories (in English) and their prototypes (in French). The WEIGHT indicates the number of words that each category contains.

THEMES	WEIGHT	Prototypes
Lived-in landscape	**40**	agréable ; agrémenté ; aider ; attaché ; vie au travers de ; bien-être ; chez ; concentré ; CUMA ; difficile ; espacé ; familial ; gens ; grand ; habitué ; impression ; incompris ; individuel ; intégration ; jeune ; liberté ; loin ; mauvais ; odeurs ; pas bon ; pas génial ; petit ; population ; qualité ; réconfortant ; s'entraider ; s'installer ; solidarité ; solitude ; sympa ; tranquille ; vacances ; vide; vivant
pattern and layout	**35**	10 15 ha ; 2 - 3 ha ; agrandissement ; carré ; cases du ; damier ; disposition ; emplacement ; est ouest ; forme ; grand ; gros ; maillage ; morcelé ; morcellement ; orienté ; parcellaire ; parcelle ; parsemé ; patchwork ; petit ; plus ; regroupé ; regroupement ; regrouper ; SAU par exploitation ; structure ; structuré ; suivent ; superficie ; surface ; taille ; terrain ; terre ; terres
crop type	**26**	à paille ; d'antan ; blé ; blé d'hiver ; blé dur ; céréales ; céréales d'hiver ; céréalier ; colza ; couverts ; grandes cultures ; maïs ; oléo-protéagineux ; orge ; permanent ; riz ; rizicole ; riziculture ; rizière ; sec ; sorgho ; tournesol ; type ; végétal ; végétaux ; vigne
esthetics	**25**	banal ; bardé bois ; beau ; beauté ; blanc ; charme ; couleur ; en accord ; fade ; formé par ; homogène ; intégration ; jaune ; joli ; lumière ; monotone ; pas trop mal ; propre ; pureté ; reflet ; riche ; tableau ; uniforme ; verdure ; vert
topography	**24**	accidenté ; bas ; bourrelet ; collinaire ; collines ; coteaux ; creux ; descend ; escarpé ; montagne ; monte ; pente ; plaine ; plan ; plat ; platitude ; relief ; terrain ; topographie ; tortueux ; vallée ; vallon ; vallonnement ; vallonné
territorial identity, culture, tradition	**23**	artisanal ; Beauce ; breton ; Camargue ; canton ; comminges ; culture ; délimité ; Gers ; Landal ; Landes ; le coin ; normand ; notre ; nous ; particulier ; pays ; poitevin ; région ; restreint ; terre ; tradition ; typique
profession, agricultural practices	**22**	abandon ; aisé ; amélioration ; amélioré ; avantageux ; bureau ; difficile ; difficulté ; facile ; méthode ; métier ; passion ; plus ; procédure ; rien ; rotation ; temps ; tout ; travail ; travaillable ; travailler ; vocation
agriculture	**21**	agricole ; agriculture ; ail ; amont ; aval ; coopérative ; cultivé ; cultural ; culture ; cultures ; exploitation ; ferme ; légumier ; maraichage ; melon ; nu ; polyculture ; représentativité ; représenté ; structure ; utilisé
evaluative judgement	**21**	95 pour cent ; à la marge ; assez ; beaucoup de ; densité ; développé ; dominant ; faible ; moins ; multitude de ; nombreuses ; peu ; peu de ; plus ; plutôt ; prédominance ; richesse ; très ; trop ; un peu ; une personne pour trois
wild nature, biodiversity	**20**	biodiversité ; canard ; chevreuil ; écosystème ; enganes ; espèce ; faune ; flamands roses ; flore ; friche ; gibier ; lande ; nature ; naturel ; oiseau ; roseau ; roseaux ; sansouïre ; sauvage ; végétation
planning, maintenance	**19**	aménagé ; aménagement ; assainissement ; dessiné ; entretenu ; entretien ; façonné ; maintient ; mis en place ; modification ; modifié ; nivelé ; plantation ; remembrement ; replantation ; replanter ; restructuration ; restructuré ; transformation
human activity	**18**	actif ; activité ; artificialisation ; artificiel ; comment ; création ; déviation ; dynamique ; emploi ; entreprise ; homme ; humain ; par accident ; par l'homme ; pluriactif ; société ; tout ; utilisation

diversity, contrasts	**17**	contraste ; de tout ; différence ; différent ; différente ; diverse ; diversification ; diversifié ; diversité ; entre ; hétérogène ; mixte ; moitié ; nuances ; ou ; plus ; varié
evolution	**16**	années 68 ; années 80 ; augmentation ; baisse ; change ; changement ; en fonction ; évolution ; gros ; moins ; moins en moins ; mutation ; nouveau ; photographie ; plus de ; stabilité
water management	**16**	amener ; canal ; digue ; domestiqué ; en eau ; évacuer ; fossé ; gérer ; hydraulique ; inondé ; irrigation ; irrigué ; maitrise ; réseau ; résolution ; roubine
soil quality	**16**	basses ; calcaire ; cultivable ; haut ; pauvre ; pierre ; qualité ; réserve utile ; riche ; salé ; sel ; sol ; terre ; terrefort ; terres ; texture
water	**15**	Aussoue ; captage ; Courance ; cours d'eau ; doux ; eau ; flotte ; Guirande ; Mignon ; Rhône ; rivière ; ruisseau ; salé ; Save ; Touch
livestock farming	**13**	agneau ; animaux ; bêtes ; bovins ; brebis ; chevaux ; chèvre ; cochon ; élevage ; mouton ; race ; taureau ; vache
view sight skyline	**12**	éloigné ; étendue ; grand ; horizon ; ligne de crête ; ouvert ; panorama ; perspective ; visage de ; vision ; voir ; vue du ciel
villages, built environment	**11**	bâti ; bourg ; château ; communes ; habitat ; hameau ; lotissement ; maison ; mas ; village ; ville
trees hedges, bocage	**10**	arbre ; arbuste ; aspect ; bocage ; bocager ; bord ; haies ; plant ; talus ; tamaris
climate, seasons	**10**	aride ; climat ; été ; hiver ; printemps ; saison ; saisons ; sec ; tempéré ; vent
environment	**9**	adaptation ; adapté ; coexistence ; conditions ; environnement ; environnemental ; résistant ; retenue ; vert
woods, forests	**8**	lisière ; taillis ; bois ; boisé ; boisement ; chaniasse ; forêt ; peupleraie
altered deteriorated	**8**	abattre ; bouleversé ; destruction ; détérioration ; détruit ; massacré ; pollution ; ravinement
places, areas	**8**	bassin de ; espace ; espaces ; milieu ; région ; sur le reste ; surface ; zone
grass, meadows	**8**	fourrager ; herbage ; herbe ; luzerne ; naturel ; prairie ; prés ; verdure
lack of	**6**	aucun ; lacunes ; non ; pas ; pas assez ; pas de
farmers, peasants	**6**	actif ; agriculteurs ; céréalier ; éleveur ; paysan ; repiqueur
agricultural decline, land abandonment	**5**	abandonné ; chômage ; clairsemé ; diminution ; reprise
extensive	**5**	autour ; dans ; dehors ; en pâture ; extensif
normative judgement	**5**	besoin de ; falloir ; harmonie ; mal ; raisonnable
agricultural production	**5**	allaitant ; laitier ; production ; utile ; viande
wetlands	**5**	étangs ; humide ; lac ; marais ; marécage
access, travel	**4**	accès ; chemins ; randonnée ; route
countryside	**4**	campagne ; champêtre ; champs ; rural
constraints, issues	**4**	contraintes ; problématique ; problème ; surcoût
public policies	**3**	administratif ; MAE ; règlementation

demonstrative pronoun	**3**	c’est ; ça ; ce que
tourism	**3**	tourisme ; touriste ; touristique
natural disaster	**2**	cataclysmes ; inondation
commerce, industry	**2**	commercial ; industriel
desert	**2**	désert ; désertique
balance	**2**	équilibre ; équilibré
intensification	**2**	augmentation ; intensif
respect, care	**2**	important ; respect

R Script to calculate thresholds and conduct the rank-frequency analyses

Below is an example of data table with the type of formatting that this script needs (Figure A1.1.):

First column: Named "**ID**" is contains the unique identification number associated with each respondent

Second column: named "**RAW**" is contains raw data, i.e. the words uttered by respondents in the form they chose (plural or singular, with typographical errors as the case may be, etc.)

Third column: named "**RANKING**" contains for each words its apparition rank, the first word uttered by a respondent get the rank 1, the second word he or she uttered gets the rank 2, etc.

Fourth column: named "**PROTOTYPICAL**", it contains prototypes associated with each uttered word, all prototypes are at the infinitive, singular, masculine form when need be. Whenever people used group of words that are not an expression (for example "beautiful forest") it must be considered as two prototypes the line must be duplicated, the rank of both prototypes is the same (the rank of the expression uttered by the respondent). When group of words are a known expression (e.g. "climate change"), it must be kept as one prototypes and put together with a dash.

Fifth column: "**THEMATICAL**" used for the categorical analysis, it contains for each uttered word the broader category it has been associated to by the researchers

Sixth, seventh and eighth columns: these are used to compare sites, social groups or inductor item (THEME).

ID	RAW	RANKING	PROTOTYPICAL	THEMATICAL	SITE	GROUP	THEME
1	hills	1	hill	topography	Camargue	F	agricultural-landscape
1	bird	2	bird	wild_fauna	Camargue	F	agricultural-landscape
1	beautiful	3	beautiful	esthetic	Camargue	F	agricultural-landscape
2	forests	1	forest	seminatural_habitat	Camargue	F	agricultural-landscape
2	cows	2	cow	livestock_farming	Camargue	F	agricultural-landscape
2	birds	3	bird	wild_fauna	Camargue	F	agricultural-landscape
2	hedge	4	hedge	seminatural_habitat	Camargue	F	agricultural-landscape
2	sunflowers	5	sunflower	cultivated_crops	Camargue	F	agricultural-landscape
3	wheat	1	wheat	cultivated_crops	Camargue	F	agricultural-landscape
3	beautiful forest	2	beautiful	esthetic	Camargue	F	agricultural-landscape
3	beautiful forest	2	forest	seminatural_habitat	Camargue	F	agricultural-landscape
4	climate change	1	climate-change	climate	Camargue	F	agricultural-landscape

Figure A1.1 Example of data file needed for the R script

 # measuring and comparing social representations
#
==========

Workspace Setting
root<-"C:/Users/" # set here the working directory

library(vegan)

CHOICE OF THE SCRIPT PARAMETERS
Theme <-"agricultural-landscape" # Indicate here the inductor word used for the free-listing task to be analysed

Site <-"Camargue" # Indicate here the name of the study site or "all" in order to pool all words from all sites together

Group <-"F" # Indicate here the social group concerned by the free-listing task (here "F" for Farmers)

opt.cat <- "PROTOTYPICAL" # Indicate here the level of categorization of the words chosen for the analysis (ex: PROTOTYPICAL for prototypical analysis or THEMATICAL for categorical analysis)

type.rang <-"apparition_Rank" # "apparition_rank"" or "importance_rank""

T <-0.1 # Indicate here the arbitrary frequency threshold (cutting point) to be used if the binomial threshold is too conservative; default data = 10%

type.freq<-"median" # choose "mean" or "median" for the crital frequency that separate central core words from the others

changes<-"yes" # when you run the script for the first time and then whenever you change the values of "theme" or "Site" or "Groupe" or "Opt.cat" set the value to "yes". In other case, set the value to "no"

Loading data
data <- read.delim(paste(root,"your-file.txt", sep=""), header=TRUE, stringsAsFactors=FALSE, sep="\t") "# indicate here the name of you text file

data.order <- data[order(data$ID),]

if (Site != "all") {t.words.cat<-data.order[which(data.order$SITE==Site & data.order$GROUP==Group & data.order$THEME==Theme),]}

if (Site == "all") {t.words.cat<-data.order[which(data.order$GROUP==Group & data.order$THEME==Theme),]}

Creating a presence - absence matrix and an order matrix
# Order matrix
key.agents <- sort(unique(t.words.cat[,"ID"]))

key.cat<-sort(unique(t.words.cat[,opt.cat]))

Ranks<-matrix(0,nrow=length(key.agents),ncol=length(key.cat))

colnames(Ranks)<-key.cat

rownames(Ranks)<-key.agents

```
for (i in 1:nrow(t.words.cat)){
   indiv<-t.words.cat[i,"ID"]
   cat<-t.words.cat[i,opt.cat]
   rg <- t.words.cat[i,"type.rang"]
  if (Ranks[indiv,cat]==0 | Ranks[indiv,cat]>rg){Ranks[indiv,cat]<-rg }
}
# Presence - absence matrix
Freq  <- (Ranks>=1)*1
```

Calculation of the Binomial Threshold

```
## Calculation of parameters for the null model
# M is the total number of different words uttered by all individuals in the study
# A is the number of individuals in the study
# mi is the number of words uttered bu the individual i
M<-ncol(Freq)
M
A<-nrow(Freq)
A
m<-as.vector(apply(Freq,1,sum))
m
## Calculation of the number of times a word j is uttered among all individuals in the study : Nobs
Nobs<-as.vector(apply(Freq, 2, sum))
Nobs.class<-sort(Nobs, decreasing=TRUE)
Nobs.class
hist.obs<-hist(Nobs,breaks=(0:(max(Nobs)+1)-0.5),plot=FALSE)
x11()
plot(hist.obs$mids,hist.obs$density,type="l",xlim=c(0,max(hist.obs$mids)),ylim=c(0,max(c(hist.obs$density)))
   ,xlab="Number of citations", ylab="Density of probability",lwd=1.5)
 legend(4,0.4,"Nobs",lty=c(1,1),lwd=c(2,2))
```

```
### Calculation of frequencies and average ranking for each word
# Frequency of citation
Freq2<-apply(Freq,2,sum)

Freq.vect<-as.vector(Freq2)

# Average Ranking
RanksNA<-ifelse(Ranks>=0,0,0)

for(i in 1:nrow(Ranks)){for (j in 1:ncol(Ranks)) {RanksNA[i,j]<-
ifelse(Ranks[i,j]==0,NA,Ranks[i,j])}}

Rg.moy<-colMeans(RanksNA, na.rm=TRUE)

Rg.moy.vect<-as.vector(Rg.moy)


# Data table = for each uttered words its frequency and average ranking
SR<-data.frame(cbind(Freq.vect,Rg.moy.vect), row.names=names(Rg.moy))

# Export data table

write.table(SR,paste(root,"SR_MF_rev_", Theme, "_", opt.cat, "_", Site,  "_", Group,".txt",
sep=""),sep="\t", dec = ",")


## The null Model
# each participant randomly picks up mi words among all words uttered by all the participants

# mi is the number of words he/she actually uttered diring the free-listing task

# when considering the raw uttered word, the null model is a "tirage sans remise" and the
probability for a word to be picked up p is p = 1 − (C_{M−1}^m / C_M^m) = m/M

# when considering the categorized data, one participant can pick up several words belonging
to the same category so that the null model become a "tirage avec remise" and p becomes p =
1 − ((M−1)/M)^m


p <- 1 - (((M-1)/M)^m) # true if opt.cat is different from "RAW" or "PROTOTYPICAL"

if (opt.cat == "RAW") {p <- m/M}

if (opt.cat == "PROTOTYPICAL") {p <- m/M}


p

length(p)

length(m)
```

```
## Calculation for each word j of its frequency of citation : Nj
Nobs<-as.vector(apply(Freq, 2, sum))

Nobs.class<-sort(Nobs, decreasing=TRUE)

Nobs.class


## Calculation of the number of words uttered by at least T people (see arbitrary threshold)
cut.arb<-T*A

M2<-length(Nobs.class[which(Nobs.class>= cut.arb)])


## Simulating what would happen if respondents randomly pick up the number of word they uttered among all word uttered. # simulate 20 000 random sampling

Z<-matrix(0,nrow=A,ncol=50000)

Nsimu<-rep(0,50000)

for (j in 1:ncol(Z))

  {

 while(sum(Z[,j])==0)

   {  for (i in 1:A)

    {

   Z[i,j]<-rbinom(1,1,p[i])

    }}

  }

Nsimu<-apply(Z,2,sum)

hist.obs<-hist(Nobs,breaks=(0:(max(Nobs)+1)-0.5),plot=FALSE)

hist.simu<-hist(Nsimu,breaks=(0:(max(Nsimu)+1)-0.5), plot=FALSE)


## graph
x11(title="Distribution de la fréquence de citation des words")

plot(hist.simu$mids,hist.simu$density,type="l",col="red",xlim=c(0,max(c(hist.simu$mids,hist.obs$mids))),ylim=c(0,max(c(hist.obs$density,hist.simu$simu)))

    ,xlab="Nombre de citations", ylab="Densité de probabilité",lwd=1.5)

lines(hist.obs$mids,hist.obs$density,type="l",lwd=1.5)

legend(4,0.4,c("Nobs","Nsimu"),lty=c(1,1),col=c("black","red"),lwd=c(2,2))
```

```
## Binomial Threshold (BT)
# The BT is a frequency threshold that represent : the lowest frequency of citation above which a word is unlikely (with a type I error rate alpha=0.05) to have been randomly picked up by respondents.


Nclass<-sort(Nsimu)

alpha<-0.05 # choose the type I error rate

cut.binom<-rep(0,M2)

for(i in M2:1)

{

cut.binom[M2-i+1]<-Nclass[(ceiling(50000*(1-(alpha/i))))]

}

cut.binom

x11(title="distribution observée de la fréquence de citation vs cut binomial et cut arbitraire")

plot(Nobs.class, ylim=c(0,max(c(Nobs[which(Nobs>=cut.arb)],cut.binom))))

lines(cut.binom,col="red2")

abline(h=cut.arb,col="blue")

legend(20,10, legend=c("cut binomial",paste("cut", T*100,"%"), col= c("red2","blue"), border="black", lwd=1))


# Critical Frequency (CF) = mean frequency of all word that reached the BT or above
words.cut.binom <- subset(SR, Freq.vect >= cut.binom[1])

CF<-mean(words.cut.binom$Freq.vect)

CF

# Arbitrary threshold
cut.arb<-T*A

# Take the minimum number between the Binomial Threshold (BT) and the arbitrary threshold (AT)
cut <- min(cut.binom, cut.arb)

# Median or Mean Frequency (MF) / NB: if BT = BT then MF = CF (Critical Frequency)
if (type.freq == "mean") {MF<-mean (words.cut$Freq.vect)}

if (type.freq == "median") {MF<-median (words.cut$Freq.vect)}

MF

# General Mean Rank
RMG<-mean(words.cut[,2])

RMG
```

Graphical outputs

```
# Keep only words uttered by C % of the people in the sample. C = cut = the appropriate threshold (arbitrary or binomial threshold)
SR2<-subset (SR,Freq.vect>= cut)

# Figure
x11()

plot(SR2[,1]~SR2[,2],

    xlim=c(min(SR2[,2])-1,max(SR2[,2])+0.7),

    ylim=c(min(SR2[,1])-1.5,max(SR2[,1])+0.5),

    xlab="Mean apparition rank",

    ylab="Frequency of citation",col="white",

    main= paste(Theme, Site, opt.cat, sep=" "))

abline(v=RMG,col="red")

abline(h=MF,col="blue")

abline(h=CF,col="darkgreen")

abline(h=cut[1],lty="dotted")

abline(h=cut.binom[1],lty="dotdash")


text(jitter(SR2[,2],2),jitter(SR2[,1],2), labels=rownames(SR2), cex = 1.3)

text(max(words.cut[,2])+0.5,MF+0.1,labels="MF",col="blue", cex = 0.8)

text(max(words.cut[,2])+0.5,CF+0.1,labels="CF",col="darkgreen", cex = 0.8)

text(max(words.cut[,2])+0.5,cut-0.1,labels="cut 10%",col="grey", cex = 0.8)

text(max(words.cut[,2])+0.5,cut.binom[1]-0.1,labels="binomial threshold", col="grey",cex = 0.8)

text(RMG-0.2,max(SR2[,1])+0.3,labels="RMG",col="red", cex = 0.8)
```

Measuring consensus level

```
# Binomial Threshold
cut.binom[1]

# General Mean Rank
RMG

# Median of Mean Frequency (depending on what was chosen earlier)
MF

# Number of Hapax
nb.Hapax<-length(Nobs.class[which(Nobs.class<2)])
```

```
nb.Hapax

# Number of consensual words
Consensus<-length(Nobs.class[which(Nobs.class>=cut.binom[1])])

Consensus

# Total number of words
M
```

Literature cited

Salès-Wuillemin, E., Morlot, R., Fontaine, A., Pullin, W., Galand, C., Talon, D., Minary-Dohen, P. 2011. Evolution of nurses' social representations of hospital hygiene: From training to practice. *European Review of Applied Psychology* 61(1) : 51–63. https://doi.org/10.1016/j.erap.2010.06.001

Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6(2):65–70.

Moran, M.D. 2003. Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos* 100(2):403–405.

Vergès, P. 1992. L'évocation de l'argent: une méthode pour la définition du noyau central d'une représentation. *Bulletin de Psychology* 65(405):203-209.