

Genre et Identité: Pour vivre heureux, vivons normés?

Pascaline Vincent

Table of contents

Introduction	3
1 Economie et Identité: quelles interactions?	4
1.1 Utilité et Identité	4
1.2 Modèle avec Identité	4
1.3 Identité et choix des individus	5
1.4 Identité et Performances	6
2 Mesurer l'identité: une impossibilité?	8
2.1 L'approche dichotomique: Hommes/Femmes	8
2.2 Les approches de mesure continue	9
2.2.1 Des échelles de mesure	9
2.2.2 Les indices composites	10
3 Notre Mesure: Le Cultural Gender Normativity Index (CGNI)	12
4 Construction de l'indice	13
4.1 Pratiques culturelles et Genre ?	13
4.1.1 Statistiques descriptives de la base de données	14
4.2 Construction de notre indice	19
4.3 Description de l'indice	27
4.4 Robustesse	29
4.5 Comparaisons avec d'autres mesures ?	31
5 Variables socio-économiques et identité de genre	33
5.1 Distance à la norme	34
5.2 Distance aux normes et satisfaction	38
6 Conclusion	49
References	51

Introduction

Avec ses cheveux courts lors de son élection, miss France 2023 a défrayé la chronique.

Au vue des réactions épidermiques, la dunkerquoise Eve Gilles semblait transgresser là une norme de genre instituée.

Avec cet exemple qui peut sembler trivial, nous souhaitons aborder la question fondamentale de l'identité, et plus spécifiquement, de l'identité de genre.

L'identité peut être définie comme le sentiment que l'on a de soi ("A person's sense of self") , c'est une notion multiple, protéiforme, difficile à appréhender et plus encore: à mesurer.

Mais pourquoi les économistes devraient-ils s'intéresser à cette variable identité?

Comment et pourquoi intègrent-ils cette notion dans leurs modèles?

Si la nécessité de proposer des modèles intégrant cette variable nous semble pertinente, vient alors la question de la difficile mesure de ce concept fondamental.

Enfin, quelles répercussions socio-économiques peut-on observer à travers la mise en lien de cette mesure de l'identité et des variables économiques?

1 Economie et Identité: quelles interactions?

L'économie de l'identité est un courant récent de la discipline.

Les travaux fondateurs d' Akerlof et Kranton (George A. Akerlof and Kranton (2000b)) ont mis en lumière la nécessité d'intégrer dans les modèles économiques traditionnels cette notion fondamentale qu'est l'identité.

1.1 Utilité et Identité

Pour un économiste, le concept d'utilité est central. Il s'agit d'une mesure de la satisfaction individuelle.

Un individu, par exemple un consommateur, réalise ses choix en fonction du niveau de satisfaction qu'il peut retirer. C'est son objectif, modélisé sous forme de fonction.

Dans les modèles traditionnels, l'utilité dépend de quelques variables objectives, qui ont l'intérêt d'être mesurables.

Par exemple, pour un consommateur, $U = F(x, y)$, où x et y sont les quantités de biens consommées.

L'apport d'Akerlof et Kranton, est d'indiquer que les choix des individus sont influencés par le sentiment qu'ils ont d'eux même, leur identité.

Enrichir les modèles traditionnels de cette nouvelle variable permettrait de mieux comprendre ce qui parfois échappe à la théorie classique.

1.2 Modèle avec Identité

Reprenons ici le modèle proposé par Akerlof et Kranton,

la fonction d'utilité d'un individu est la suivante:

$$U_j = U_j(a_j, a_j I_j)$$

où :

- a_j représente les actions de l'individu j

- a_j I_j capture les interactions entre les actions de j et son identité I_j

Cette fonction d'identité I_j dépend elle même de plusieurs facteurs:

$$I_j = I_j(a_j, a_{-j}, c_j, E_j, P)$$

où :

- a_j représente les actions de l'individu j ,
- a_{-j} représente les actions des autres individus,
- c_j est la catégorie sociale à laquelle appartient j ,
- E_j sont les caractéristiques données de j ,
- P sont les prescriptions (ou normes sociales) associées à la catégorie c_j .

Ce modèle formalise cette tension qui peut exister pour les individus entre se conformer ou non aux normes de sa catégorie sociale assignée, en fonction de la distance qui peut exister entre ses propres caractéristiques et les prescriptions assignées.

Il suppose aussi l'existence possible de sanctions de la part des pairs si les normes ne sont pas suivies (moqueries, mise à l'écart ou même violence)

Par exemple, dans leur papier (George A. Akerlof and Kranton, n.d.) complètent les théories traditionnelles de l'économie de l'éducation en expliquant le plus ou moins grand investissement scolaire des étudiants en fonction de leurs caractéristiques identitaires (les sportifs, les intellos et les rebelles)

1.3 Identité et choix des individus

L'économie étudie comment les agents font leur choix (de consommation, de production...), ces choix dépendent de l'identité des individus.

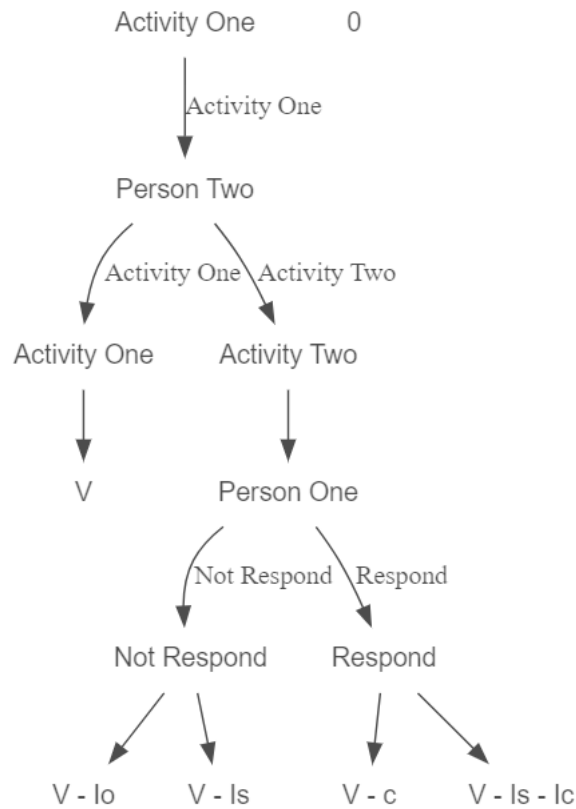
En effet, le sens que l'on a de soi, influence nos décisions, et ces décisions influencent notre identité en retour.

Dans un premier temps, la catégorie sociale à laquelle j'appartiens peut influencer mes préférences Chen and Li (2009), mes comportements pro-sociaux Charness, Cobo-Reyes, and Jiménez (2014) et donc mes décisions.

Dans un second temps, l'individu arbitre entre ses propres caractéristiques et les prescriptions de sa catégorie sociale (ou les prescriptions supposées).

Vais-je me conformer à ce qui est attendu? Mes propres caractéristiques sont-elles plus ou moins proches des normes en vigueur? Si je ne me conforme pas, quelle sera la réaction de mes pairs?

Cet arbitrage et cette possibilité pour autrui de répondre à mes actions a été représenté sous forme d'arbre de décision par George A. Akerlof and Kranton (2000a)



Cette friction entre adhérer ou non aux normes de ma catégorie sociale a conduit la recherche des économistes de l'identité vers les questions de conflits inter-groupes Austen-Smith and Fryer (2005), Chakravarty et al. (2015), Fryer, Fryer, and Torelli (2010), Bénabou and Tirole (2004).

On s'interroge sur les normes, l'adhésion ou non à ces dernières et le rôle que peuvent jouer les politiques publiques sur ces prescriptions.

Mais l'identité joue également un rôle fondamental sur les performances individuelles.

1.4 Identité et Performances

Dans une étude particulièrement intéressante, Shih, Pittinsky, and Trahan (2006) ont montré comment l'activation d'une identité particulière, ou tout du moins comment les stéréotypes

associés à cette identité, pouvaient avoir un impact sur les performances individuelles.

En réalisant une étude auprès d'étudiantes asiatiques, les chercheurs ont proposé l'expérience suivante:

Un questionnaire était fourni aux participantes avant de réaliser un test de mathématiques.

Une partie des participantes répondaient à un questionnaire dont les questions portaient sur leur identité asiatique, pour les autres participantes le questionnaire activait leur identité féminine, un groupe de contrôle était également implémenté.

Ce simple questionnaire en préambule a eu des effets notables sur les performances aux tests de mathématiques réalisés ensuite, puisque en moyenne, les étudiantes dont l'identité asiatique avait été mise en évidence performaient significativement mieux aux tests de mathématiques que leurs homologues dont l'identité féminine avait été activée.

Cette étude plaide donc pour la nécessité de s'interroger sur la notion d'identité, dès lors que cette variable peut significativement avoir des répercussions sur les résultats socio-économiques des individus.

Cependant, les critiques apportées à ce nouveau courant sont notamment que ce concept fondamental d'identité est trop flou, comment intégrer dans les modèles ou bien encore dans les études empiriques cette dimension si polymorphe?

La question de la mesure de l'identité se pose alors...

2 Mesurer l'identité: une impossibilité?

Si nous acceptons l'idée qu'il est pertinent pour les économistes d'intégrer dans leurs modèles la variable Identité, alors la question de sa mesure est fondamentale.

L'identité est multiple, dans ce document, nous nous focalisons sur l'identité de genre des individus.

Une des raisons pour lesquelles nous portons notre attention sur cette dimension est que l'identité de genre nous semble être l'une des plus importantes dans la construction individuelle.

Avant même la naissance d'un enfant, la question fondamentale de son sexe biologique est omniprésente: *c'est une fille ou un garçon?*

L'organisation de la société est centrée autour de cette distinction genrée, les inégalités persistantes entre hommes et femmes justifient également notre intérêt pour cette dimension de l'identité.

Les questionnements sur la non-binarité influencent également notre réflexion, peut-on proposer une mesure de l'identité de genre qui tienne compte d'un continuum entre deux pôles (masculin et féminin), permettant ainsi de mesurer une distance à ces extremum?

Nous verrons donc comment est abordée la mesure de l'identité de genre dans la littérature existante, puis proposerons une mesure continue de cette identité, avec une approche empirique basée sur les pratiques culturelles (genrées) des français.

2.1 L'approche dichotomique: Hommes/Femmes

Lorsque la dimension du genre veut être prise en compte dans les études économiques, bien souvent on se réfère au sexe biologique des individus: homme/femme.

Cette approche binaire est notamment liée à la commodité des données.

En effet, dans les bases de données, cette information figure presque toujours, elle permet d'étudier et de mettre en évidence des différences, souvent des inégalités, entre les deux sexes biologiques.

Cependant, certains individus ne se retrouvent pas dans cette dichotomie, alors la possibilité d'intégrer une troisième option dans les questionnaires est parfois envisagée.

Mais là encore, une approche catégorielle peine à saisir ce qui nous semble fondamental dans une étude empirique de l'identité: la distance aux Prescriptions.

2.2 Les approches de mesure continue

Afin de prendre en compte la diversité des positions individuelles le long d'un axe Masculin/Féminin, des études ont proposé une approche continue de mesure de l'identité de genre. Nous les distinguons de la façon suivante: certaines études reposent sur un positionnement individuel de l'individu sur une échelle de mesure.

D'autres études proposent la construction d'indices composites pour mesurer la dimension continue du genre.

2.2.1 Des échelles de mesure

Les travaux proposant des mesures continues du genre basées sur des échelles de gradation reposent sur les déclarations individuelles des enquêtés, qui se positionnent le long d'un ou plusieurs axes.

Par exemple, Magliozzi, Saperstein, and Westbrook (2016) propose aux enquêtés de se situer sur plusieurs échelles (comment je me perçois: de très masculin à pas du tout masculin, et de très féminin à pas du tout féminin), mais aussi comment les autres me perçoivent.

First-order gender scale

In general, how do you see yourself? Please answer on both scales below.

	Not at all	1	2	3	4	5	Very
Feminine	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Masculine	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Third-order gender scale

In general, how do most people see you? Please answer on both scales below.

	Not at all	1	2	3	4	5	Very
Feminine	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Masculine	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Sex at birth

What sex were you assigned at birth?
(For example, on your birth certificate.)

☐ Female

☐ Male

☐ Intersex

Categorical gender identification

What is your current gender?

☐ Woman

☐ Man

☐ Transgender

☐ A gender not listed here (please specify)

En France, l'étude Virage a permis de poser des questions similaires, les résultats de cette étude ont fait l'objet d'un papier rédigé par Trachman (2022a).

Il est intéressant de voir qu'en effet, l'identité de genre est faite de nuances, la plupart des individus interrogés ne s'identifient pas aux pôles extrêmes de leur sexe biologique (23,3% des femmes interrogées s'identifient comme très féminines; 30,6% des hommes interrogés s'identifient comme très masculins).

Peu d'individus dévient des normes (1% des femmes interrogées se disent très masculines, 2,3% des hommes interrogés se disent très féminins.)

Ces résultats mettent en lumière la complexité de l'identité de genre, et peut-être confirment la nécessité d'aller au delà du binaire dans les mesures proposées.

Cette approche par gradation a l'avantage de demander directement aux individus le sentiment qu'ils ont d'eux même, et donc de tendre vers la définition même de l'identité.

L'inconvénient repose sur l'accessibilité des données, elle suppose d'intégrer ces questions dans les questionnaires et ces informations sont donc peu disponibles mais liées à des enquêtes bien spécifiques.

Un autre biais est peut-être également que cette approche repose sur la subjectivité des personnes interrogées, elles peuvent être influencées par le biais de la désirabilité sociale (je réponds avec la peur d'être jugé si je réponds *mal*, je réponds ce qui me semble être conforme aux normes).

Il peut donc être intéressant de combiner ces approches avec d'autres mesures plus indirectes mais, peut-être, moins subjectives.

2.2.2 Les indices composites

Dans la famille des mesures continues de l'identité de genre, certains chercheurs ont construit leurs propres indices de mesures continues.

2.2.2.1 Avec Définitions *a priori* des dimensions genrées

Parmi les indices composites, un outil très utilisé est le Bem Sex Role Inventory (Bem (1974)).

Il s'agit de créer un indice composite à partir de réponses des enquêtés qui se classent par rapport à différents items.

Ces items sont identifiés au préalable comme plutôt masculin (ex:agressif), plutôt féminin (ex: lunatique) ou neutres (ex: naïf).

Cette approche permet de capturer une distance aux normes, cependant elle suppose de définir au préalable ce qui relève de l'ordre du masculin et du féminin.

2.2.2.2 Sans Définitions *a priori*

Une autre approche consiste à construire un indice composite à partir de variables identifiées statistiquement comme genrées, c'est à dire déduire de l'analyse statistique les pratiques ou comportements plus ou moins masculins ou féminins (dans la mesure où ils sont statistiquement significativement différenciés selon les sexes biologiques) et de construire un indice en fonction des poids que représentent chaque variable sur un axe masculin/féminin.

Cette approche est notamment utilisée en médecine (Pelletier, Ditto, and Pilote (2015).)

Nous avons suivi cette dernière approche pour notre étude, en nous référant notamment aux travaux de Cipriani et al. (n.d.) qui offre un guide pratique pour la réalisation d'un tel indicateur.

3 Notre Mesure: Le Cultural Gender Normativity Index (CGNI)

4 Construction de l'indice

L'indice que nous proposons repose donc sur la dernière approche présentée:

Un indice de mesure continue de l'identité de genre, construit comme un indice composite à partir de dimensions non définies *a priori* comme genrées.

Mais alors, quelles variables choisir pour représenter ces dimensions du genre?

Notre choix s'est porté sur les pratiques culturelles des français (leurs loisirs culturels) car ces pratiques sont en effet particulièrement genrées (différenciées selon les sexes biologiques), il nous paraissait pertinent de s'appuyer sur ces dernières pour construire notre indice.

4.1 Pratiques culturelles et Genre ?

En sociologie, la question de la culture et du Genre fait l'objet de travaux ayant montré combien les pratiques culturelles sont différenciées chez les hommes et les femmes Octobre (2008) , et ce, dès l'enfance.

Cette différenciation nous permet d'envisager qu'il y ait des pratiques plus ou moins féminines ou masculines, dans la mesure où en moyenne elles sont plus pratiquées par des hommes ou par des femmes.

Cela signifie qu'il existe des normes genrées dans la pratique ou non d'une activité culturelle: le tricot est essentiellement féminin, la chasse est une activité plutôt pratiquée par les hommes.

Ces exemples sont tirés de l'analyse de nos données, en effet, nous avons utilisé la base de données *Enquête sur les pratiques culturelles des Français*, 2018.

Cette base de données comprend des informations sur les pratiques culturelles des français (9234 individus interrogés) , ainsi que des données socio-démographiques et porte également une question qui va nous intéresser sur le degré de satisfaction en termes de temps libre.

Cette dernière variable ("Vous arrive-t-il d'avoir le sentiment de manquer de temps libre pour faire tout ce dont vous avez envie?") a retenu notre attention car elle pourrait être une mesure de l'utilité (satisfaction) de l'individu.

4.1.1 Statistiques descriptives de la base de données

```
library(foreign)
library(questionr)
library(ggplot2)
library(tidyverse)
library(ggmosaic)
library(GGally)
library(dataMaid)
library(dplyr)
library(GDAtools)
library(FactoMineR)
library(gtsummary)
library(factoextra)
library(gtsummary)
library(kableExtra)
library(RColorBrewer)
library(FactoMineR)
library(xtable)
library(explor)
```

```
data<-read.csv2("pc18_quetelet_octobre2023.csv")
```

```
data$Sex <- factor(data$SEXE,
                   levels = c(1, 2),
                   labels = c("Men", "Women"))
my_data_frame <- data |> dplyr::rename(
  Knitting = A1001 ,
  Cards_games = A1002,
  Gambling = A1003 ,
  Cooking = A1004 ,
  DIY = A1005 ,
  Vegetable_garden = A1006 ,
  Ornamental_garden = A1007,
  Fishing_hunting = A1008 ,
  Collection = A1009 ,
  Vehicle_custom = A1010 ,
  No_Amateur=A1011,
  Making_music = A1901 ,
  Diary = A1902 ,
  Writing = A1903,
```

```

Painting = A1904,
Montage = A1905 ,
Circus = A1906 ,
Pottery = A1907 ,
Theater = A1908 ,
Drawing = A1909 ,
Dancing = A1910,
Photography = A1911 ,
Genealogy = A1912 ,
Science = A1913 ,
None = A1914 ,
Video_games = B1 ,
TV = C1 ,
Radio = E1 ,
Library = F1 ,
Museums = H112,
Internet = I4 ,
Concert = G2413 )

my_data_frame$Video_games <- ifelse(my_data_frame$Video_games == 1, 1, 0)
my_data_frame$TV <- ifelse(my_data_frame$TV == 5, 0, 1)
my_data_frame$Radio <- ifelse(my_data_frame$Radio == 5, 0, 1)
my_data_frame$Library<- ifelse(my_data_frame$Radio == 1, 1, 0)
my_data_frame$Museums<- ifelse(my_data_frame$Museums == 1, 0, 1)
my_data_frame$Internet<- ifelse(my_data_frame$Internet == 5, 0, 1)
my_data_frame$Concert<- ifelse(my_data_frame$Concert == 1, 0, 1)

my_data_frame <- my_data_frame %>%
  mutate(satisfaction = case_when(
    A2 %in% 1 ~ "Low",      # 1 à 4 -> Low
    A2 %in% 2 ~ "Medium",   # 5 à 7 -> Medium
    A2 %in% 3 ~ "High",     # 8 à 10 -> High
    TRUE ~ NA_character_
  ))
my_data_frame <- my_data_frame %>%
  mutate(Income = case_when(
    CRITREVENU %in% 1:4 ~ "Low",      # 1 à 4 -> Low
    CRITREVENU %in% 5:7 ~ "Medium",   # 5 à 7 -> Medium
    CRITREVENU %in% 8:10 ~ "High",    # 8 à 10 -> High
    TRUE ~ NA_character_
  ))

```

```

))

my_data_frame <- my_data_frame %>%
  mutate(Health = case_when(
    A15 %in% 1:2 ~ "Good",      # 1 à 4 -> Low
    A15 %in% 3 ~ "Medium",     # 5 à 7 -> Medium
    A15 %in% 4:5 ~ "Bad",      # 8 à 10 -> High
    TRUE ~ NA_character_
  ))

my_data_frame <- my_data_frame %>%
  mutate(Couple = case_when(
    VITENCOUPLE %in% 1:2 ~ "Yes",
    VITENCOUPLE %in% 3 ~ "No",

    TRUE ~ NA_character_
  ))

quartiles <- quantile(data$AGE, probs = c(0.25, 0.5, 0.75), na.rm = TRUE)

my_data_frame$age_group <- cut(
  my_data_frame$AGE,
  breaks = 4, # Automatically divide into 4 slices
  labels = c("[15-38[", "[38-54[", "[54-67[", "[67-97[", # Labels optionnels
  include.lowest = TRUE
)

stat_des <- my_data_frame |>
  tbl_summary(
    include = c("AGE", "Income", "Couple", "CLASSIF", "satisfaction", "Health"),
    by = "Sex",
    statistic = list(
      all_continuous() ~ "{min} - {max}"
    )
  ) |>
  add_overall(last = TRUE) |>
  add_p(
    test.args = list(
      all_continuous() ~ list(simulate.p.value = TRUE),
      all_categorical() ~ list(simulate.p.value = TRUE)
    )
  )

```



```
)  
  
stat_des
```

Lecture: Parmi les individus de sexe masculin, 59% sont en couple. La différence avec les femmes est significative ($p < 0,001$)

```
table <- my_data_frame |>  
tbl_summary(  
  include = c( "Knitting" ,  
               "Cards_games",  
               "Gambling" ,  
               "Cooking" ,  
               "DIY" ,  
               "Vegetable_garden" ,  
               "Ornamental_garden",  
               "Fishing_hunting" ,  
               "Collection" ,  
               "Vehicle_custom",  
               "Making_music" ,  
               "Diary" ,  
               "Writing" ,  
               "Painting",  
               "Montage" ,  
               "Circus" ,  
               "Pottery" ,  
               "Theater" ,  
               "Drawing" ,  
               "Dancing",  
               "Photography" ,  
               "Genealogy" ,  
               "Science" ,  
               "None" ,  
               "No_Amateur",  
               "Video_games" ,  
               "TV" ,  
               "Radio" ,  
               "Library" ,  
               "Museums",  
               "Internet",  
               "Concert"),  
  
  by = "Sex"
```

Characteristic	Men N = 4,162 ¹	Women N = 5,072 ¹	Overall N = 9,234 ¹	p-value ²
AGE	15 - 95	15 - 97	15 - 97	0.042
Income				<0.001
High	1,432 (39%)	1,464 (33%)	2,896 (36%)	
Low	762 (21%)	1,284 (29%)	2,046 (25%)	
Medium	1,476 (40%)	1,644 (37%)	3,120 (39%)	
Unknown	492	680	1,172	
Couple	2,444 (59%)	2,567 (51%)	5,011 (54%)	<0.001
Unknown	4	9	13	
CLASSIF				<0.001
1	123 (8.7%)	63 (4.5%)	186 (6.6%)	
2	454 (32%)	143 (10%)	597 (21%)	
3	190 (13%)	92 (6.6%)	282 (10%)	
4	157 (11%)	126 (9.1%)	283 (10%)	
5	278 (20%)	236 (17%)	514 (18%)	
6	194 (14%)	703 (51%)	897 (32%)	
7	13 (0.9%)	7 (0.5%)	20 (0.7%)	
8	11 (0.8%)	18 (1.3%)	29 (1.0%)	
9	1 (<0.1%)	1 (<0.1%)	2 (<0.1%)	
Unknown	2,741	3,683	6,424	
satisfaction				<0.001
High	1,461 (35%)	1,656 (33%)	3,117 (34%)	
Low	1,488 (36%)	2,015 (40%)	3,503 (38%)	
Medium	1,207 (29%)	1,398 (28%)	2,605 (28%)	
Unknown	6	3	9	
Health				<0.001
Bad	322 (7.8%)	461 (9.1%)	783 (8.5%)	
Good	2,984 (72%)	3,426 (68%)	6,410 (70%)	
Medium	841 (20%)	1,157 (23%)	1,998 (22%)	
Unknown	15	28	43	

¹Min - Max; n (%)

²Wilcoxon rank sum test; Pearson's Chi-squared test with simulated p-value (based on 2000 replicates); Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)

```
) |>
add_overall(last = TRUE) |>
add_p()

table
```

Lecture: Parmi les hommes, 1.8% déclarent pratiquer le tricot, contre 27% des femmes. Cette activité est pratiquée par 16% des répondants.

Les p-value <10% indiquent que les différences de participation aux pratiques culturelles sont significativement différentes selon le sexe biologique du répondant.

i Note

Nous retiendrons pour la construction de notre indice les pratiques culturelles suivantes: "Knitting", "Cards_games", "Gambling", "Cooking", "DIY", "Vegetable_garden", "Fishing_hunting", "Collection", "Vehicle_custom", "Making_music", "Diary", "Writing", "Painting", "Montage", "Pottery", "Theater", "Drawing", "Dancing", "Photography", "Genealogy", "Science", "None", "Video_games", "Library", "Concert"

4.2 Construction de notre indice

Afin de construire notre indice d'identité de genre, nous suivons la méthodologie proposée par Cipriani et al. (n.d.) et réalisons une Analyse des Correspondances Multiples (ACM) sur nos variables "pratiques culturelles".

```
pratiques_cols_1 <- c( "Knitting" ,
                      "Cards_games",
                      "Gambling" ,
                      "Cooking" ,
                      "DIY" ,
                      "Vegetable_garden" ,
                      "Fishing_hunting" ,
                      "Collection" ,
                      "Vehicle_custom",
                      "Making_music" ,
                      "Diary" ,
                      "Writing" ,
                      "Painting",
                      "Montage" ,
```

Characteristic	Men N = 4,162 ¹	Women N = 5,072 ¹	Overall N = 9,234 ¹	p-value ²
Knitting	73 (1.8%)	1,370 (27%)	1,443 (16%)	<0.001
Cards_games	1,899 (46%)	2,764 (54%)	4,663 (50%)	<0.001
Gambling	990 (24%)	970 (19%)	1,960 (21%)	<0.001
Cooking	1,685 (40%)	3,473 (68%)	5,158 (56%)	<0.001
DIY	2,667 (64%)	2,283 (45%)	4,950 (54%)	<0.001
Vegetable_garden	1,335 (32%)	1,245 (25%)	2,580 (28%)	<0.001
Ornamental_garden	1,793 (43%)	2,198 (43%)	3,991 (43%)	0.8
Fishing_hunting	722 (17%)	212 (4.2%)	934 (10%)	<0.001
Collection	395 (9.5%)	281 (5.5%)	676 (7.3%)	<0.001
Vehicle_custom	264 (6.3%)	60 (1.2%)	324 (3.5%)	<0.001
Making_music	1,312 (32%)	1,830 (36%)	3,142 (34%)	<0.001
Diary	285 (6.8%)	1,206 (24%)	1,491 (16%)	<0.001
Writing	410 (9.9%)	746 (15%)	1,156 (13%)	<0.001
Painting	667 (16%)	1,303 (26%)	1,970 (21%)	<0.001
Montage	843 (20%)	586 (12%)	1,429 (15%)	<0.001
Circus	116 (2.8%)	179 (3.5%)	295 (3.2%)	0.044
Pottery	264 (6.3%)	705 (14%)	969 (10%)	<0.001
Theater	483 (12%)	800 (16%)	1,283 (14%)	<0.001
Drawing	864 (21%)	1,285 (25%)	2,149 (23%)	<0.001
Dancing	410 (9.9%)	1,782 (35%)	2,192 (24%)	<0.001
Photography	1,175 (28%)	1,176 (23%)	2,351 (25%)	<0.001
Genealogy	513 (12%)	575 (11%)	1,088 (12%)	0.14
Science	611 (15%)	469 (9.2%)	1,080 (12%)	<0.001
None	1,394 (33%)	1,261 (25%)	2,655 (29%)	<0.001
No_Amateur	276 (6.6%)	359 (7.1%)	635 (6.9%)	0.4
Video_games	1,760 (42%)	1,827 (36%)	3,587 (39%)	<0.001
TV	3,878 (93%)	4,806 (95%)	8,684 (94%)	0.001
Radio	3,522 (85%)	4,186 (83%)	7,708 (83%)	0.007
Library	3,522 (85%)	4,186 (83%)	7,708 (83%)	0.007
Museums	4,101 (99%)	5,009 (99%)	9,110 (99%)	0.4
Internet	3,459 (83%)	4,185 (83%)	7,644 (83%)	0.4
Concert	3,344 (80%)	4,234 (83%)	7,578 (82%)	<0.001

¹n (%)

²Pearson's Chi-squared test

```

        "Pottery" ,
        "Theater" ,
        "Drawing" ,
        "Dancing",
        "Photography" ,
        "Genealogy" ,
        "Science" ,
        "None" ,
        "Video_games" ,
        "Library" ,
        "Concert"
      )

#MCA

# Add Sex Column to the selection
cols_of_interest_1 <- c("Sex", "AGE", pratiques_cols_1)

# Build a new dataframe with these columns
data_pratiques <- my_data_frame[, cols_of_interest_1]
data_pratiques$AGE <- cut(data_pratiques$AGE,
                          breaks = quantile(data_pratiques$AGE, probs = seq(0, 1, 0.25), na.rm = TRUE),
                          include.lowest = TRUE)
ra_data <- na.omit(data_pratiques)

cols_to_factor <- c( "Knitting" ,
                     "Cards_games",
                     "Gambling" ,
                     "Cooking" ,
                     "DIY" ,
                     "Vegetable_garden" ,
                     "Fishing_hunting" ,
                     "Collection" ,
                     "Vehicle_custom",
                     "Making_music" ,
                     "Diary" ,
                     "Writing" ,
                     "Painting",
                     "Montage" ,
                     "Pottery" ,
                     "Theater" ,

```

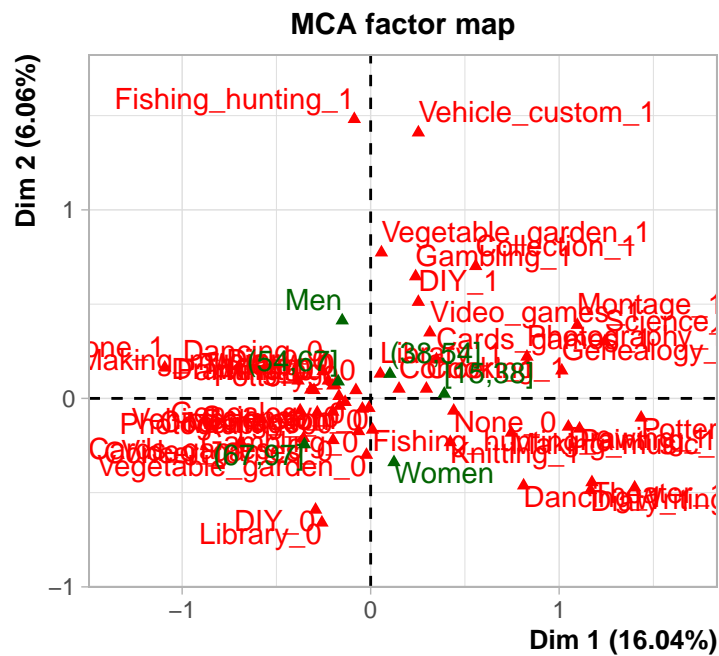
```

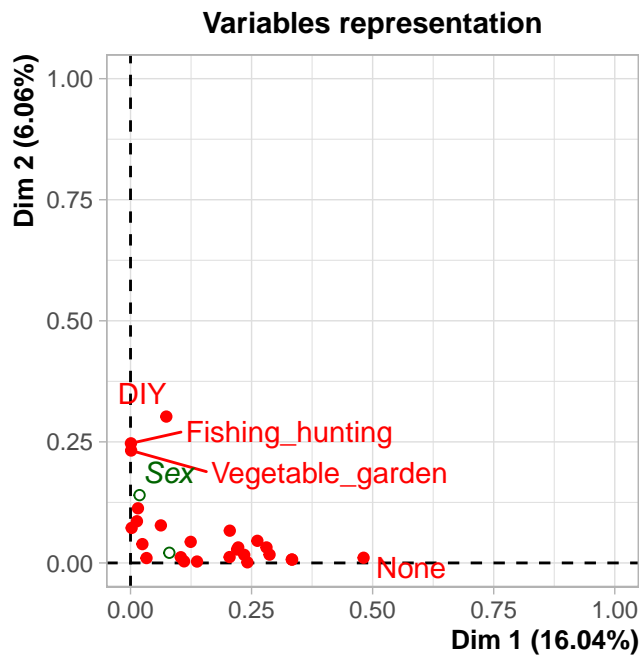
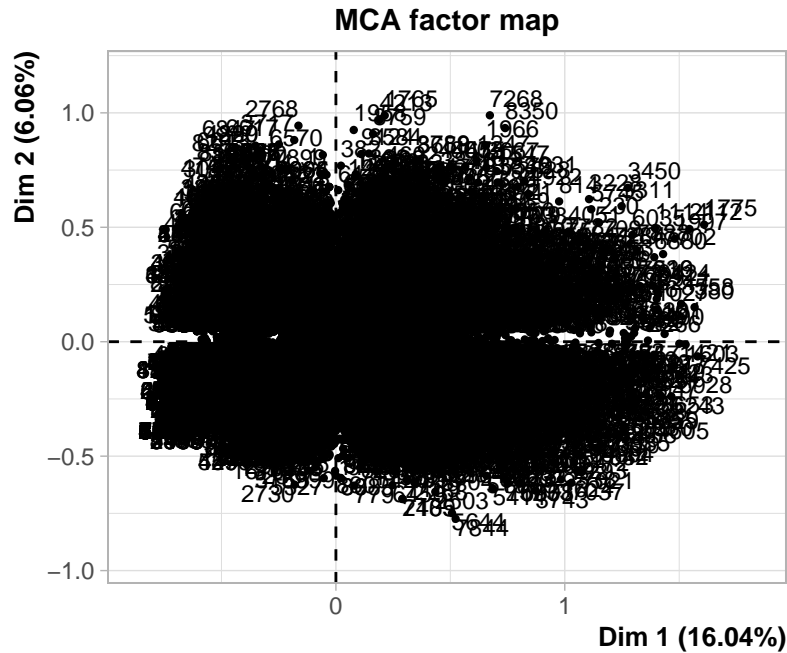
    "Drawing" ,
    "Dancing",
    "Photography" ,
    "Genealogy" ,
    "Science" ,
    "None" ,
    "Video_games" ,
    "Library" ,
    "Concert")

# apply as.factor to these columns
ra_data[cols_to_factor] <- lapply(ra_data[cols_to_factor], as.factor)

# running MCA with FactoMineR
acm2_fm <- ra_data |>
  FactoMineR::MCA(
    ncp = Inf,
    graph = TRUE,
    quali.sup = 1:2
  )

```





On remarque que la variable supplémentaire “Sexe” correspond à la dimension 2 de mon ACM. Cette dimension explique 6,06% de la variance totale.

Une analyse plus poussée avec le package Explor nous permet de mesurer l’association de la variable supplémentaire Sexe avec cette dimension (dim2), en effet, l’ η^2 est de 0,14 , ce qui

peut sembler peu mais indique bien que notre variable supplémentaire est liée à cet axe.

Pour construire notre indice, nous utiliserons donc les coordonnées des variables pratiques culturelles le long de cet axe 2.

```
# Extract modality names
modalites_names <- rownames(acm2_fm$var$coord)
```

```
# Check modality names
head(modalites_names)
```

```
[1] "Knitting_0"      "Knitting_1"      "Cards_games_0"  "Cards_games_1"
[5] "Gambling_0"      "Gambling_1"
```

```
# Extract coordinates for dimension 2
coord_dim2_modalites <- acm2_fm$var$coord[, 2]
```

```
# Create a table associating the modalities and their coordinates in dimension 2
modalites_coord <- data.frame(Modalite = modalites_names, Coord_Dim2 = coord_dim2_modalites)
```

```
# Keep only the two necessary columns
modalites_coord_selected <- modalites_coord[, c("Modalite", "Coord_Dim2")]

print(modalites_coord_selected)
```

	Modalite	Coord_Dim2
Knitting_0	Knitting_0	0.04314603
Knitting_1	Knitting_1	-0.23295268
Cards_games_0	Cards_games_0	-0.21072530
Cards_games_1	Cards_games_1	0.20656774
Gambling_0	Gambling_0	-0.17433328
Gambling_1	Gambling_1	0.64698995
Cooking_0	Cooking_0	-0.06432635
Cooking_1	Cooking_1	0.05083253
DIY_0	DIY_0	-0.59097111
DIY_1	DIY_1	0.51145863
Vegetable_garden_0	Vegetable_garden_0	-0.29999865
Vegetable_garden_1	Vegetable_garden_1	0.77371744
Fishing_hunting_0	Fishing_hunting_0	-0.16671349
Fishing_hunting_1	Fishing_hunting_1	1.48150102

Collection_0	Collection_0	-0.05539977
Collection_1	Collection_1	0.70134797
Vehicle_custom_0	Vehicle_custom_0	-0.05128080
Vehicle_custom_1	Vehicle_custom_1	1.41022187
Making_music_0	Making_music_0	0.09403755
Making_music_1	Making_music_1	-0.18232870
Diary_0	Diary_0	0.09378209
Diary_1	Diary_1	-0.48702531
Writing_0	Writing_0	0.06770001
Writing_1	Writing_1	-0.47308017
Painting_0	Painting_0	0.04361213
Painting_1	Painting_1	-0.16081142
Montage_0	Montage_0	-0.07125251
Montage_1	Montage_1	0.38917132
Pottery_0	Pottery_0	0.01201096
Pottery_1	Pottery_1	-0.10244644
Theater_0	Theater_0	0.07187594
Theater_1	Theater_1	-0.44542916
Drawing_0	Drawing_0	0.04570585
Drawing_1	Drawing_1	-0.15068681
Dancing_0	Dancing_0	0.14396501
Dancing_1	Dancing_1	-0.46250072
Photography_0	Photography_0	-0.07539973
Photography_1	Photography_1	0.22074706
Genealogy_0	Genealogy_0	-0.01967874
Genealogy_1	Genealogy_1	0.14733732
Science_0	Science_0	-0.04008198
Science_1	Science_1	0.30261895
None_0	None_0	-0.06572453
None_1	None_1	0.16286315
Video_games_0	Video_games_0	-0.22185919
Video_games_1	Video_games_1	0.34927205
Library_0	Library_0	-0.65936864
Library_1	Library_1	0.13053925
Concert_0	Concert_0	-0.23244943
Concert_1	Concert_1	0.05079655

Ce tableau indique les poids utilisés pour la construction de notre indice.

```
# Initialize a vector to store the index of each individual
ra_data$indice_culturel <- 0
```

```

# Browse each individual
for (i in 1:nrow(ra_data)) {

  # Initialize individual's index to 0
  indice_individu <- 0

  # Browse each practice column (columns 3 to 27)
  for (pratique in 3:27) {

    # Retrieve the individual's response for this practice (0 or 1)
    reponse <- ra_data[i, pratique]

    # If the answer is 1, add the coordinate of the corresponding modality to the index.
    if (reponse == 1) {

      # Create the modality name (e.g. "knitting_1" or "knitting_0")
      nom_modalite_1 <- paste0(names(ra_data)[pratique], "_1")
      nom_modalite_0 <- paste0(names(ra_data)[pratique], "_0")

      # Find the coordinate associated with the corresponding modality
      if (nom_modalite_1 %in% modalites_coord$Modalite) {
        indice_individu <- indice_individu + modalites_coord$Coord_Dim2[modalites_coord$Modalite]
      }
      if (nom_modalite_0 %in% modalites_coord$Modalite) {
        indice_individu <- indice_individu + modalites_coord$Coord_Dim2[modalites_coord$Modalite]
      }
    }
  }

  # Assign the calculated index to the individual
  ra_data$indice_culturel[i] <- indice_individu
}

####Normalisation

# Calculate minimum and maximum index values
min_indice <- min(ra_data$indice_culturel, na.rm = TRUE)
max_indice <- max(ra_data$indice_culturel, na.rm = TRUE)

# Normalize index
ra_data$indice_culturel_normalise <- (ra_data$indice_culturel - min_indice) / (max_indice - min_indice)

# Check results

```

```
head(ra_data[, c("indice_culturel", "indice_culturel_normalise")])
```

	indice_culturel	indice_culturel_normalise
1	2.8024423	0.8624142
2	0.5427446	0.5073268
3	-0.4557477	0.3504244
4	-0.3429253	0.3681533
5	0.0409659	0.4284777
6	-0.2367887	0.3848315

Notre indice est donc construit de la façon suivante:

$$I_{1j} = \sum_{k=1}^Z w_{1k} \cdot X_{kj}$$

Dans cette expression, I_{1j} désigne l'indice de l'individu j , tandis que w_{1k} représente le poids associé à chaque variable culturelle X_{kj} . La somme englobe toutes les variables culturelles Z , ce qui nous permet de saisir l'engagement culturel global de l'individu.

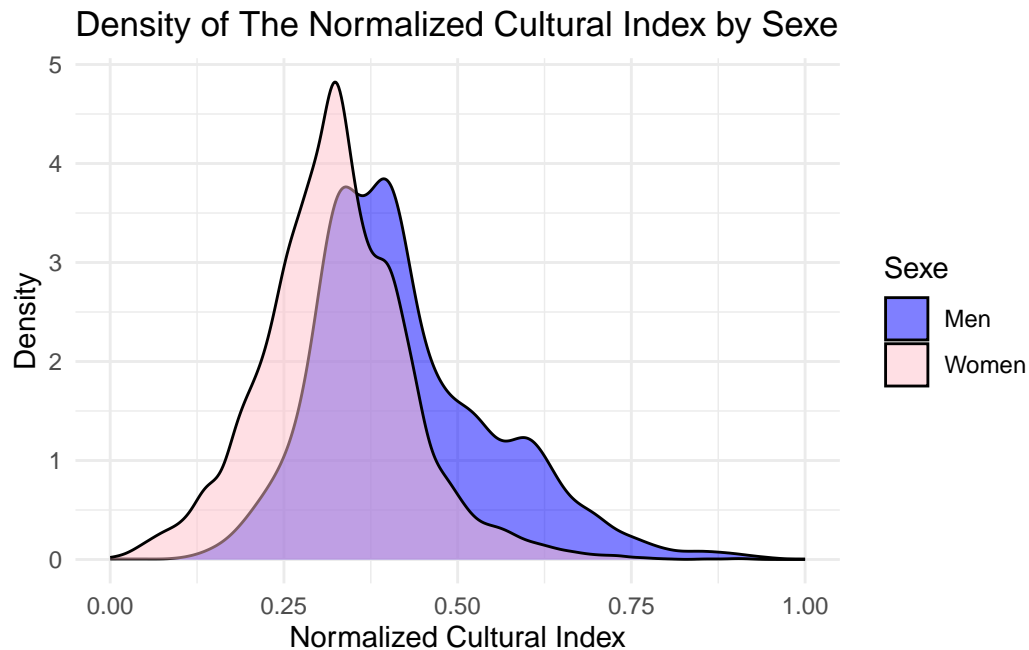
4.3 Description de l'indice

```
my_data_frame$identity<-ra_data$indice_culturel_normalise
my_data_frame$indice<-ra_data$indice_culturel
ggplot(my_data_frame, aes(x = identity, fill = Sex)) +
  geom_density(alpha = 0.5) +
  scale_fill_manual(values = c("blue", "pink")) +
  labs(title = "Density of The Normalized Cultural Index by Sexe",
       x = "Normalized Cultural Index",
       y = "Density",
       fill = "Sexe") +
  theme_minimal()
```

Characteristic	Men N = 4,162 ¹	Women N = 5,072 ¹	Overall N = 9,234 ¹	p-value ²
identity	0.01 - 1.00	0.00 - 0.91	0.00 - 1.00	<0.001

¹Min - Max

²Wilcoxon rank sum test



```
stat_des_indice<-my_data_frame%>%
  tbl_summary(include=c("identity"),by = "Sex",
              statistic = list(
                all_continuous() ~ "{min} - {max}"
              )) %>%
  add_overall(last = TRUE) %>%
  add_p()
stat_des_indice
```

L'indice normalisé est compris entre 0 et 1. Plus il est proche de zéro plus les individus sont proches de la norme féminine (en termes de pratiques culturelles).

L'indice est significativement différent selon le sexe biologique des interrogés.

```
reg<- lm(identity~Sex, my_data_frame)
summary(reg)
```

Call:

```
lm(formula = identity ~ Sex, data = my_data_frame)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.41933	-0.07875	-0.01093	0.06694	0.58782

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.424436	0.001815	233.79	<2e-16 ***
SexWomen	-0.099669	0.002450	-40.69	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1171 on 9232 degrees of freedom

Multiple R-squared: 0.1521, Adjusted R-squared: 0.152

F-statistic: 1656 on 1 and 9232 DF, p-value: < 2.2e-16

Le sexe biologique est très significatif, les valeurs plus faibles de l'indice sont associées au sexe féminin.

4.4 Robustesse

```
data$identity<- my_data_frame$identity
my_data_frame$identity<-data$identity
# List of cultural activities
cultural_activities <- c(
  "Knitting" ,
  "Cards_games",
  "Gambling" ,
  "Cooking" ,
  "DIY" ,
  "Vegetable_garden" ,
  "Fishing_hunting" ,
  "Collection" ,
```

```

        "Vehicle_custom",
        "Making_music" ,
        "Diary" ,
        "Writing" ,
        "Painting",
        "Montage" ,
        "Pottery" ,
        "Theater" ,
        "Drawing" ,
        "Dancing",
        "Photography" ,
        "Genealogy" ,
        "Science" ,
        "None" ,
        "Video_games" ,
        "Library" ,
        "Concert"
    )

# Create a data frame to store the results
result_table <- data.frame(Activity = character(0), Accuracy = numeric(0))

for (activity in cultural_activities) {
  # Perform a Probit regression for the current cultural activity
  model_formula <- as.formula(paste(activity, "~ identity"))
  model <- glm(model_formula, data = my_data_frame, family = binomial(link = "probit"))

  # Calculate predictions
  predicted <- ifelse(predict(model, type = "response") >= 0.5, 1, 0)

  # Calculate the accuracy
  correct_predictions <- sum(predicted == my_data_frame[[activity]])
  total_predictions <- nrow(my_data_frame)
  accuracy <- (correct_predictions / total_predictions) * 100

  # Add the result to the data frame
  result_table <- rbind(result_table, data.frame(Activity = activity, Accuracy = accuracy))
}
result_table

```

	Activity	Accuracy
1	Knitting	84.37297

2	Cards_games	51.00715
3	Gambling	79.18562
4	Cooking	56.55187
5	DIY	51.25623
6	Vegetable_garden	72.89365
7	Fishing_hunting	93.95712
8	Collection	92.54927
9	Vehicle_custom	96.85943
10	Making_music	68.16114
11	Diary	85.85662
12	Writing	88.16331
13	Painting	78.76327
14	Montage	84.52458
15	Pottery	89.50617
16	Theater	86.53888
17	Drawing	76.73814
18	Dancing	79.66212
19	Photography	74.53974
20	Genealogy	88.21746
21	Science	88.30409
22	None	70.19710
23	Video_games	62.06411
24	Library	83.20338
25	Concert	81.92549

Nous réalisons une série de régression Probit sur les différentes pratiques culturelles, avec pour unique régresseur notre indice d'identité, afin de mesurer le pouvoir prédictif (ou accuracy) de ce dernier.

4.5 Comparaisons avec d'autres mesures ?

Comparons l'indice obtenu avec d'autres mesures, et plus particulièrement avec les indices échelle d'identité.

Pour cela, nous allons diviser notre indice en 7 catégories, allant du plus féminin au plus masculin.

```
my_data_frame$identity_scale <- cut(
  my_data_frame$identity,
  breaks = 7, # Automatically divide into 7 slices
  labels = c("Very Feminine", "1", "2", "3", "4", "5", "Very Masculine"),
```

Characteristic	Men N = 4,162 ¹	Women N = 5,072 ¹	p-value ²
identity_scale			<0.001
Very Feminine	9 (0.2%)	218 (4.3%)	
1	403 (9.7%)	1,506 (30%)	
2	2,139 (51%)	2,669 (53%)	
3	985 (24%)	576 (11%)	
4	509 (12%)	90 (1.8%)	
5	96 (2.3%)	10 (0.2%)	
Very Masculine	21 (0.5%)	3 (<0.1%)	

¹n (%)

²Pearson's Chi-squared test

```

include.lowest = TRUE
)
table1 <-
  my_data_frame |>
  tbl_summary(include = c(identity_scale),
    by=Sex ,)>
  add_p()
table1

```

Ce tableau est à mettre en perspective avec les données de Trachman (2022a).

Tout d'abord remarquons que, comme dans l'enquête de Trachman (2022a), les individus "hors norme" sont peu nombreux (d'après notre indice, moins de 0,1% des femmes ont des pratiques culturelles très masculines; 0,2% des hommes ont des pratiques très féminines), de même que les individus dont les pratiques culturelles seraient complètement conformes à leur sexe biologique ne sont pas la majorité (4,3% des femmes sont classées comme très féminines, 0,5% des hommes comme très masculins.)

Cela plaide encore une fois pour l'intérêt d'une mesure continue, nous allons ensuite explorer plus en détails les variables socio-démographiques qui peuvent influencer notre indice (revenu, âge, catégorie socio-professionnelle ...)

Dans cette partie, nous allons analyser les implications socio-économiques de l'identité de genre.

Pour commencer, nous allons regarder les différents profils des individus selon leur indice d'identité.

Puis nous analyserons les liens entre distance à la norme et l'utilité des individus (leur degré de satisfaction.)

5 Variables socio-économiques et identité de genre

```
stat_des <- my_data_frame |>
tbl_summary(
  include = c("AGE", "Income", "Couple", "CLASSIF"), by = "identity_scale") |>

add_overall(last = TRUE) |>
add_p(
  test.args = list(
    all_continuous() ~ list(simulate.p.value = TRUE), all_categorical() ~ list
  )
)

stat_des
```

Lecture: Parmi les individus catégorisés comme très féminins selon notre indice, 44% ont des revenus élevés. Les variables étudiées sont significativement différentes selon les degrés d'identité.

Il semble, d'après cette première table descriptive, que l'âge, les niveaux de revenus ainsi que les professions (CLASSIF) jouent un rôle dans l'identité de genre. Tout comme dans l'étude de Trachman (2022b), les facteurs socio-économiques influencent les variations du genre.

CLASSIF	
1	Manoeuvre ou ouvrier spécialisé
2	Ouvrier qualifié ou hautement qualifié/ technicien(ne) d'atelier
3	Technicien(ne)
4	Agent de maîtrise, maîtrise administrative ou commerciale, VRP (non cadre)
5	Ingénieur, Cadre

CLASSIF	
6	Employé(e) de bureau, Employé(e) de commerce, Personnel de services
7	Directeur général, Adjoint direct
8	NSP
9	REF

On peut regarder si les effets sont différents selon le sexe biologique:

```
stat_des <- my_data_frame |>
  tbl_strata(
    strata = Sex, # Stratification par sexe
    .tbl_fun = ~ .x |>
      tbl_summary(
        include = c("AGE", "Income", "Couple", "CLASSIF"),
        by = "identity_scale" # Stratification supplémentaire par 'identity_scale'
      ) |>
    add_overall(last = TRUE) |>
    add_p(
      test.args = list(
        all_continuous() ~ list(simulate.p.value = TRUE),
        all_categorical() ~ list(simulate.p.value = TRUE)
      )
    )
  )

stat_des
```

Lecture: Parmi les hommes classés comme très féminins, 27% appartiennent à la catégorie socio-professionnelle 5 (ingénieur, cadre)

5.1 Distance à la norme

Nous l'évoquions au début de ce document, ce qui nous intéresse en proposant une mesure continue des variations du genre c'est de pouvoir analyser les distances prises avec les normes de son groupe de référence.

Pour cela, nous allons construire la variable "Distance à la norme"

Characteristic	Very Feminine N = 227¹	1 N = 1,909¹	2 N = 4,808¹	3 N = 1,561¹	4 N = 590¹
AGE	44 (32, 57)	50 (35, 65)	56 (40, 69)	55 (37, 68)	51 (35, 65)
Income					
High	88 (44%)	710 (42%)	1,428 (34%)	452 (33%)	175 (33%)
Low	48 (24%)	382 (23%)	1,118 (27%)	369 (27%)	108 (20%)
Medium	65 (32%)	592 (35%)	1,631 (39%)	536 (39%)	245 (46%)
Unknown	26	225	631	204	71
Couple	109 (48%)	1,013 (53%)	2,594 (54%)	842 (54%)	373 (62%)
Unknown	2	4	5	1	1
CLASSIF					
1	0 (0%)	18 (3.2%)	95 (6.8%)	48 (10%)	16 (6.6%)
2	6 (8.0%)	61 (11%)	278 (20%)	132 (28%)	98 (40%)
3	3 (4.0%)	36 (6.4%)	148 (11%)	56 (12%)	33 (14%)
4	8 (11%)	59 (11%)	140 (10.0%)	47 (9.8%)	22 (9.1%)
5	26 (35%)	149 (27%)	229 (16%)	80 (17%)	26 (11%)
6	29 (39%)	224 (40%)	484 (34%)	110 (23%)	47 (19%)
7	1 (1.3%)	6 (1.1%)	11 (0.8%)	1 (0.2%)	1 (0.4%)
8	2 (2.7%)	6 (1.1%)	16 (1.1%)	4 (0.8%)	0 (0%)
9	0 (0%)	0 (0%)	2 (0.1%)	0 (0%)	0 (0%)
Unknown	152	1,350	3,405	1,083	356

¹Median (Q1, Q3); n (%)

²Pearson's Chi-squared test with simulated p-value (based on 2000 replicates); Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)

Characteristic	Men				
	Very Feminine N = 9 ¹	1 N = 403 ¹	2 N = 2,139 ¹	3 N = 985 ¹	4 N = 509 ¹
AGE	50 (36, 59)	54 (39, 66)	54 (40, 68)	52 (35, 66)	52 (36, 65)
Income					
High	4 (44%)	160 (44%)	762 (40%)	316 (36%)	153 (34%)
Low	4 (44%)	66 (18%)	393 (21%)	192 (22%)	88 (20%)
Medium	1 (11%)	134 (37%)	733 (39%)	359 (41%)	203 (46%)
Unknown	0	43	251	118	65
Couple	4 (44%)	224 (56%)	1,271 (59%)	561 (57%)	314 (62%)
Unknown	0	1	2	0	1
CLASSIF					
1	0 (0%)	6 (5.2%)	57 (7.9%)	37 (11%)	16 (7.7%)
2	0 (0%)	24 (21%)	214 (30%)	109 (33%)	87 (42%)
3	0 (0%)	15 (13%)	98 (14%)	42 (13%)	30 (14%)
4	0 (0%)	15 (13%)	76 (11%)	37 (11%)	22 (11%)
5	1 (100%)	31 (27%)	151 (21%)	66 (20%)	25 (12%)
6	0 (0%)	21 (18%)	104 (15%)	41 (12%)	26 (13%)
7	0 (0%)	2 (1.7%)	9 (1.3%)	1 (0.3%)	1 (0.5%)
8	0 (0%)	1 (0.9%)	7 (1.0%)	2 (0.6%)	0 (0%)
9	0 (0%)	0 (0%)	1 (0.1%)	0 (0%)	0 (0%)
Unknown	8	288	1,422	650	302

¹Median (Q1, Q3); n (%)

²Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)

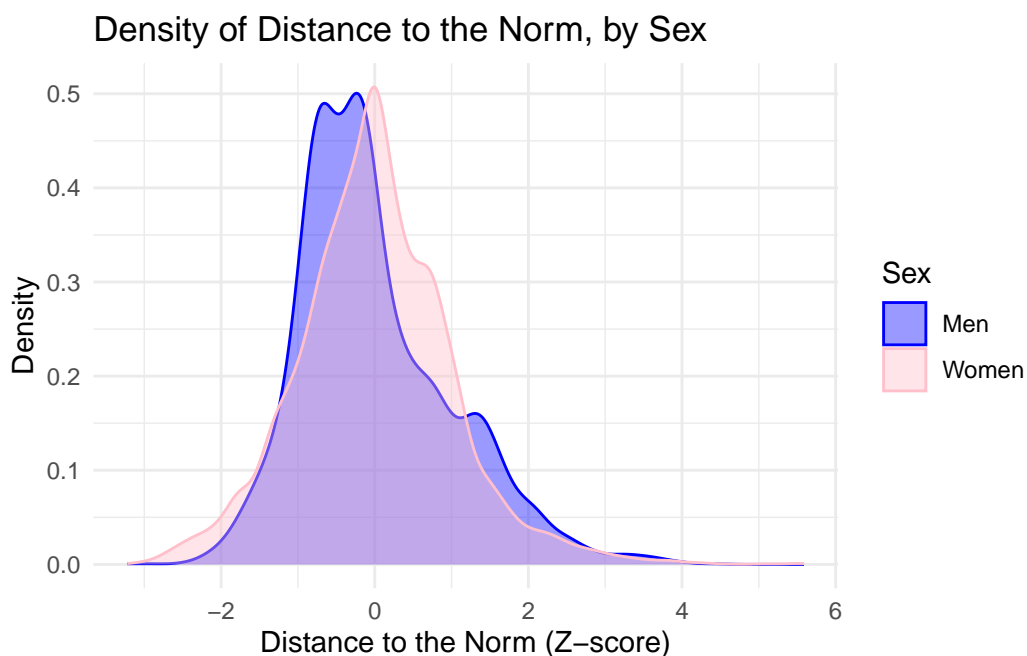
```

mean_gender <- tapply(my_data_frame$identity, my_data_frame$Sex, mean)
sd_gender <- tapply(my_data_frame$identity, my_data_frame$Sex, sd)

my_data_frame$distance<- (my_data_frame$identity - mean_gender[my_data_frame$Sex]) / sd_gender[my_data_frame$Sex]

ggplot(my_data_frame, aes(x = distance, color = Sex, fill = Sex)) +
  geom_density(alpha = 0.4) +
  labs(title = "Density of Distance to the Norm, by Sex",
       x = "Distance to the Norm (Z-score)", y = "Density") +
  scale_fill_manual(values = c("blue", "pink")) +
  scale_color_manual(values = c("blue", "pink")) +
  theme_minimal()

```



```

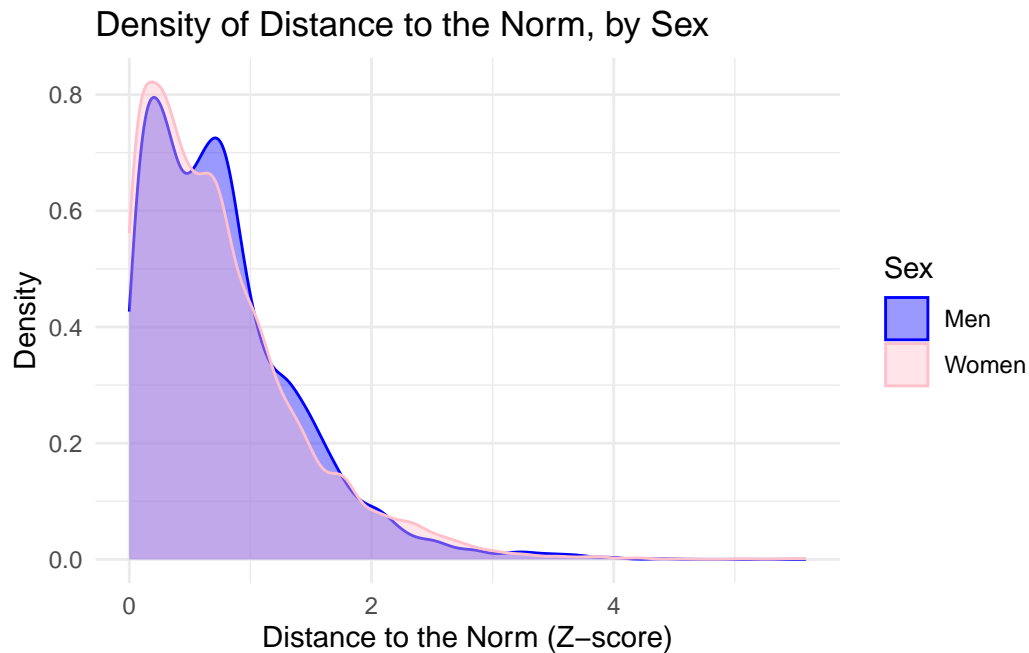
#Ou en valeur absolue:
mean_gender <- tapply(my_data_frame$identity, my_data_frame$Sex, mean)
sd_gender <- tapply(my_data_frame$identity, my_data_frame$Sex, sd)

my_data_frame$distance_abs<- abs((my_data_frame$identity - mean_gender[my_data_frame$Sex]) / sd_gender[my_data_frame$Sex])

ggplot(my_data_frame, aes(x = distance_abs, color = Sex, fill = Sex)) +
  geom_density(alpha = 0.4) +
  labs(title = "Density of Distance to the Norm, by Sex",

```

```
x = "Distance to the Norm (Z-score)", y = "Density") +
scale_fill_manual(values = c("blue", "pink")) +
scale_color_manual(values = c("blue", "pink")) +
theme_minimal()
```



```
# Catégorisation par quartiles
my_data_frame$categories <- cut(my_data_frame$distance,
```

```
breaks = quan
```

5.2 Distance aux normes et satisfaction

```
library(MASS)
my_data_frame$satisfaction<-as.factor((my_data_frame$satisfaction))
# Ordinal logistic regression model fitting
ordinal_model <- polr(satisfaction ~ distance, data = my_data_frame, Hess = TRUE)

# 1. Predict the probabilities for each category
distance_vals <- seq(min(my_data_frame$distance), max(my_data_frame$distance))
new_data <- data.frame(distance = distance_vals)
```

```

predicted_probs <- predict(ordinal_model, newdata = new_data, type = "probs")

# 3. Convert results to dataframe
predicted_probs_df <- as.data.frame(predicted_probs)
predicted_probs_df$distance <- distance_vals

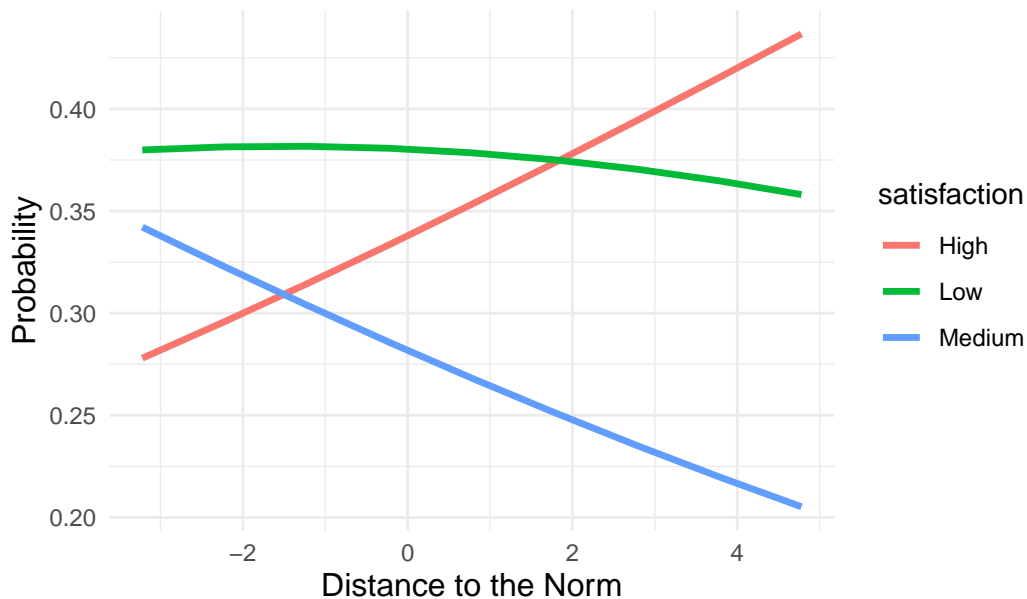
# 4. Transform data into long format for ggplot2
predicted_probs_long <- pivot_longer(predicted_probs_df,
                                     cols = -distance,
                                     names_to = "satisfaction",
                                     values_to = "probabilite")

# 5. Visualize predicted probabilities
ggplot(predicted_probs_long, aes(x = distance, y = probabilite, color = satisfaction)) +
  geom_line(size = 1.2) +
  labs(title = "Probability of satisfaction as a function of deviation from Norms",
       x = "Distance to the Norm",
       y = "Probability") +

  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
    axis.title.x = element_text(size = 12),
    axis.title.y = element_text(size = 12)
  )

```

ability of satisfaction as a function of deviation from Norms



La satisfaction est mesurée comme étant la réponse à la question “Vous arrive-t-il d’avoir le sentiment de manquer de temps libre pour faire tout ce dont vous avez envie ?”

Ce graphique semble indiquer que la satisfaction en termes de temps libre croît avec la distance à sa norme de genre.

Cet effet est-il vérifié pour les différents niveaux de revenus et le sexe biologique?

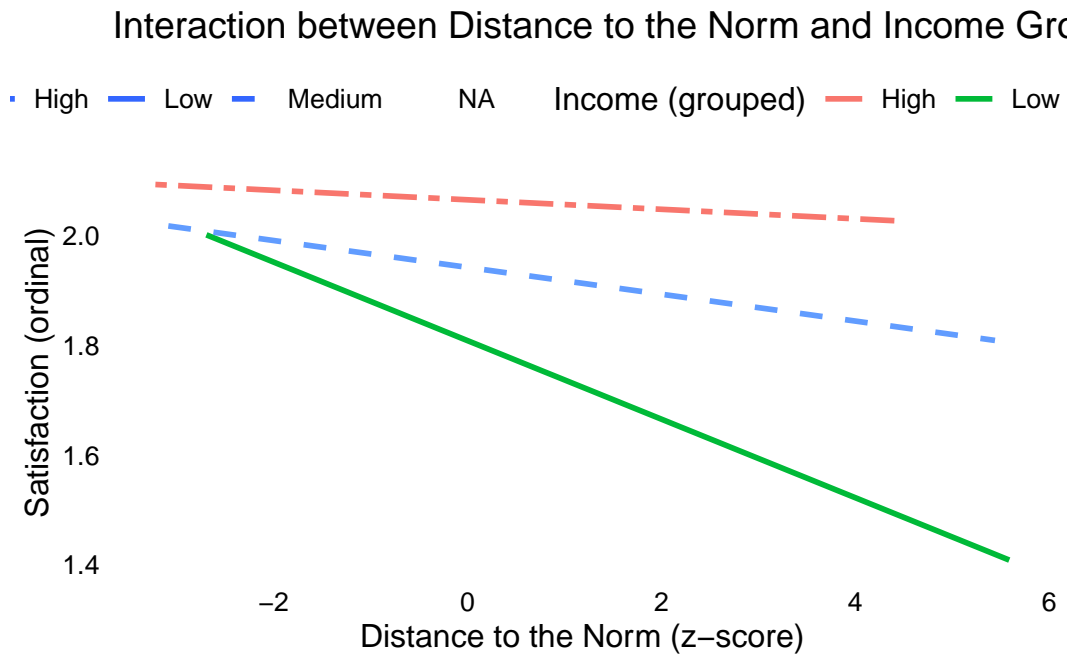
```
# Graph with only lines and no scatter points
ggplot(my_data_frame, aes(x = distance, y = as.numeric(satisfaction), color = as.factor(Income))) +
  geom_smooth(method = "lm", aes(linetype = as.factor(Income)), se = FALSE) + # Regression ci
  scale_linetype_manual(values = c("Low" = "solid", "Medium" = "dashed", "High" = "twodash"))
labs(
  title = "Interaction between Distance to the Norm and Income Groups",
  x = "Distance to the Norm (z-score)",
  y = "Satisfaction (ordinal)",
  color = "Income (grouped)",
  linetype = "Line Type"
) +
theme_minimal() + # Minimalist theme
theme(
  text = element_text(size = 12), # Text size
  panel.grid = element_blank(), # Remove gridlines
  panel.border = element_blank(), # Remove borders
```



```

plot.background = element_blank(),
legend.position = "top",
axis.title = element_text(color = "black"),
axis.text = element_text(color = "black")
)
# Remove gray background
# Place legend at top
# Axis titles in black
# Axis text in black

```



L'effet semble s'inverser pour les bas revenus, s'écarter de la norme s'accompagne d'une baisse de la satisfaction ressentie.

Qu'en est-il si on regarde ces effets par sexe?

```

base_femmes <- my_data_frame %>%
  filter(Sex == "Women")

# Creating a table without NA
femmes_categories <- base_femmes %>%
  tbl_summary(
    include = c(Income, LNAIS, Couple, satisfaction, Health, age_group),
    by = categories,
    statistic = ~ "{p}%",
    percent = "row",
    missing = "no"
  ) %>%

```

Characteristic	Feminine Norms N = 1,183 ¹	Medium N = 2,607 ¹	Masculine Norms N = 1,183 ¹
Income			
High	30%	50%	19%
Low	19%	52%	28%
Medium	22%	50%	28%
LNAIS			
1	24%	51%	25%
2	20%	53%	27%
Couple satisfaction	23%	52%	25%
High	16%	53%	31%
Low	29%	50%	21%
Medium	24%	51%	25%
Health			
Bad	16%	54%	30%
Good	26%	51%	23%
Medium	18%	52%	29%
age_group			
[15-38[30%	49%	21%
[38-54[27%	49%	24%
[54-67[19%	54%	27%
[67-97[12%	56%	31%

¹%

²Pearson's Chi-squared test

```
add_p()
femmes_categories
```

```
base_hommes <- my_data_frame %>%
  filter(Sex == "Men")

hommes_categories <- base_hommes %>%
  tbl_summary(
    include = c(Income, LNAIS, Couple, satisfaction, Health, age_group),
    by = categories,
    statistic = ~ "{p}%",
    percent = "row",
```

Characteristic	Feminine Norms N = 1,126 ¹	Medium N = 2,009 ¹	Masculine Norms N = 1,126 ¹
Income			
High	30%	48%	23%
Low	26%	50%	24%
Medium	25%	47%	27%
LNAIS			
1	26%	48%	26%
2	33%	53%	13%
Couple satisfaction	26%	49%	25%
High	26%	50%	24%
Low	28%	46%	26%
Medium	27%	49%	24%
Health			
Bad	25%	53%	22%
Good	28%	48%	25%
Medium	26%	48%	26%
age_group			
[15-38[24%	46%	30%
[38-54[27%	49%	24%
[54-67[28%	47%	25%
[67-97[31%	55%	14%

¹%

²Pearson's Chi-squared test

```

missing = "no"
) %>%
  add_p()
hommes_categories

```

```

# Fit ordinal logistic regression model including SEX
ordinal_model2 <- polr(satisfaction ~ distance * Sex, data = my_data_frame, Hess = TRUE)

# Create a sequence of values for conformity
conformity_vals <- seq(min(my_data_frame$distance), max(my_data_frame$distance))

# Create a new dataset with all combinations of conformity and SEX
new_data <- expand.grid(distance = distance_vals, Sex = c("Men", "Women"))

```

```

# Predict probabilities for each combination of conformity and SEXE
predicted_probs <- predict(ordinal_model2, newdata = new_data, type = "probs")

# Convert results to dataframe
predicted_probs_df <- as.data.frame(predicted_probs)
predicted_probs_df$distance <- new_data$distance
predicted_probs_df$Sex <- new_data$Sex

# Correctly name probability columns for satisfaction levels 1, 2 and 3
colnames(predicted_probs_df)[1:3] <- c("satisfaction_1", "satisfaction_2", "satisfaction_3")

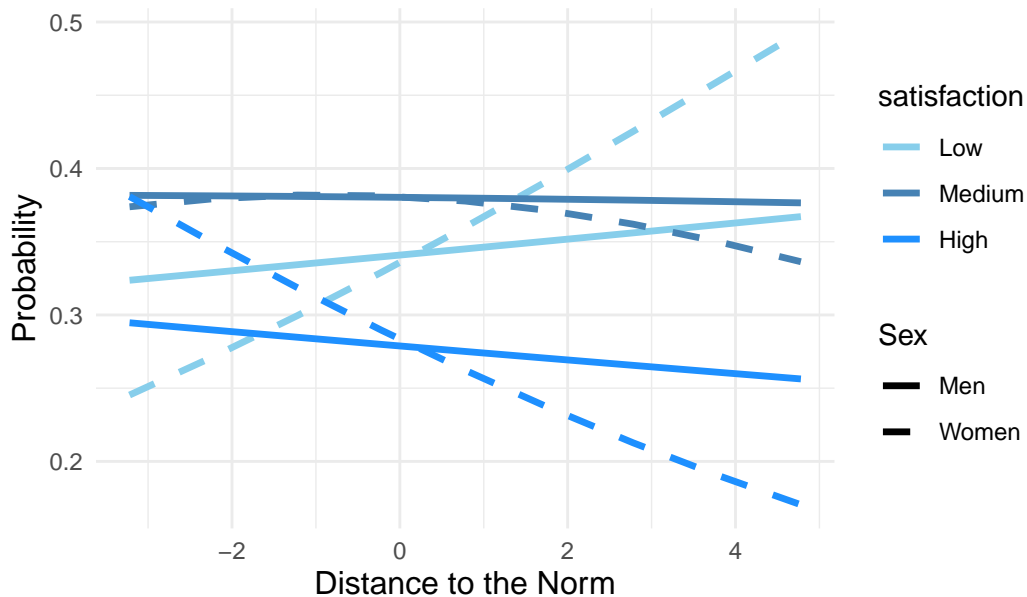
# Transformation to long format for easier viewing with ggplot2
predicted_probs_long <- pivot_longer(predicted_probs_df,
                                     cols = starts_with("satisfaction"), # Selects columns s
                                     names_to = "satisfaction",
                                     values_to = "probabilite")

# Modify names for simple labels
predicted_probs_long$satisfaction <- factor(predicted_probs_long$satisfaction,
                                           levels = c("satisfaction_1", "satisfaction_2", "satisfaction_3"),
                                           labels = c("Low", "Medium", "High"))

# Visualize predicted probabilities for each gender
ggplot(predicted_probs_long, aes(x = distance, y = probabilite, color = satisfaction, linetype = gender)) +
  geom_line(size = 1.2) +
  labs(title = "Predicted Probabilities of Satisfaction Based on Distance to the Norm",
       x = "Distance to the Norm",
       y = "Probability") +
  scale_color_manual(values = c("#87CEEB", "#4682B4", "#1E90FF")) + #
  scale_linetype_manual(values = c("solid", "dashed")) + # Differentiates between men and women
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
    axis.title.x = element_text(size = 12),
    axis.title.y = element_text(size = 12)
  )

```

d Probabilities of Satisfaction Based on Distance to the Norm



Ce graphique montre des différences significatives entre hommes et femmes, chez les femmes l'éloignement aux normes semble augmenter l'insatisfaction, cet effet n'est pas aussi net chez les hommes.

```
rego <- MASS::polr(
  satisfaction ~ Sex + Income + age_group + categories,
  data = my_data_frame
)

theme_gtsummary_language("en", decimal.mark = ",")
rego |>
  tbl_regression(
    exponentiate = TRUE,
    tidy_fun = broom.helpers::tidy_parameters
  ) |>
  bold_labels() |>
  add_global_p(keep = TRUE) |>
  as_kable_extra(format = "latex", booktabs = TRUE) |>
  kable_styling(latex_options = c("hold_position", "scale_down"), full_width = FALSE)
```

```
rego |>
  ggstats::ggcoef_table(
```

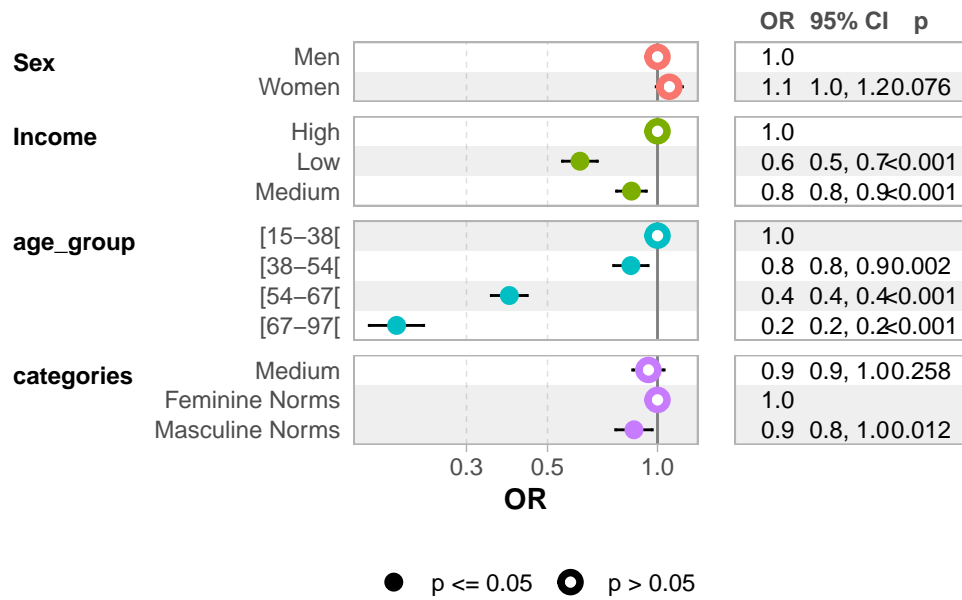
Characteristic	OR	95% CI	p-value
Sex			0,076
Men	—	—	
Women	1,08	0,99 – 1,17	0,076
Income			<0,001
High	—	—	
Low	0,61	0,55 – 0,68	<0,001
Medium	0,85	0,77 – 0,93	<0,001
age_group			<0,001
[15-38[—	—	
[38-54[0,85	0,76 – 0,94	0,002
[54-67[0,39	0,35 – 0,44	<0,001
[67-97[0,19	0,16 – 0,23	<0,001
categories			0,040
Feminine Norms	—	—	
Medium	0,94	0,86 – 1,04	0,3
Masculine Norms	0,86	0,77 – 0,97	0,012

¹ OR = Odds Ratio, CI = Confidence Interval

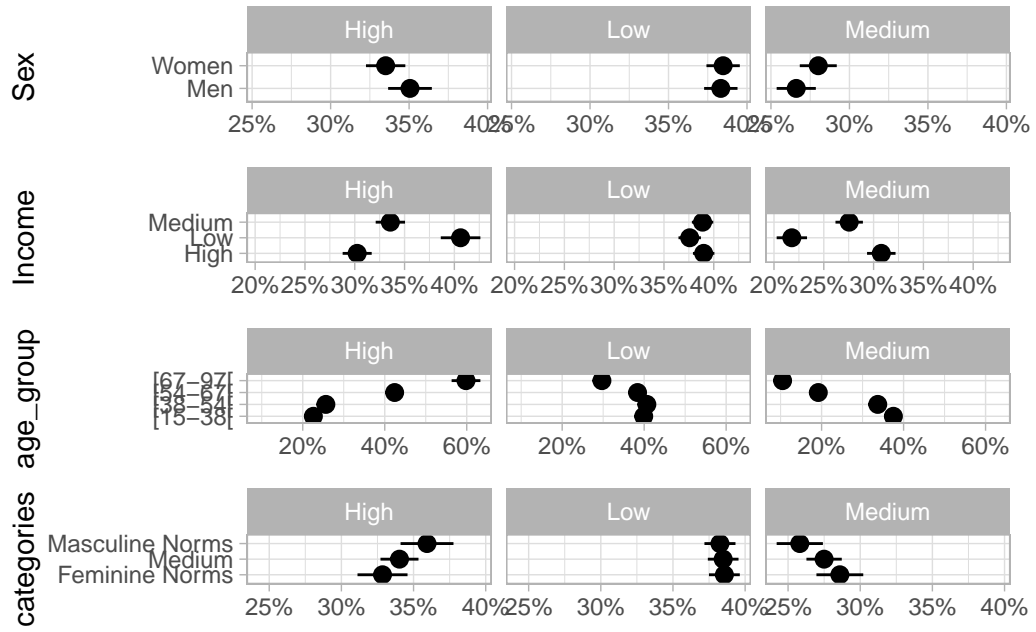
```

exponentiate = TRUE,
tidy_fun = broom.helpers::tidy_parameters
)

```



```
rego |>
  broom.helpers::plot_marginal_predictions() |>
  patchwork::wrap_plots(ncol = 1) &
  scale_y_continuous(labels = scales::percent) &
  coord_flip()
```



6 Conclusion

Nous avons pu voir l'intérêt que portent les économistes aux questions d'identité.

Champ récent de la discipline, l'économie de l'identité propose plusieurs directions de recherche: d'un point de vue théorique, elle cherche à introduire dans des modèles économiques traditionnels la variable identité, afin d'enrichir notre compréhension des phénomènes économiques.

Parce qu'elle influence les choix, l'identité doit être prise en compte dans une discipline où ces choix individuels sont un objet d'études.

L'identité influence les préférences individuelles et notamment la coopération (à la participation de biens publics par exemple), elle agit sur les performances individuelles et influence les actions des individus.

Ces derniers font face à des arbitrages entre leurs caractéristiques intrinsèques et les normes sociales (ou prescriptions) qui prévalent dans leur environnement.

Il y a donc là de nombreux questionnements sur les dynamiques de formation de l'identité et ses répercussions socio-économiques .

D'un point de vue empirique, l'une des premières difficultés auxquelles fait face le chercheur est la mesure de cette variable aux contours flous et aux multiples dimensions.

L'emploi de variables catégorielles (comme la variable binaire homme/femme) contient des limites lorsque l'objectif est de mesurer la distance aux normes d'identité.

Les données continues provenant de déclarations et positionnements individuels sur une échelle graduée (par exemple, dans quelle mesure je me perçois comme très masculin/féminin), répond en partie aux problèmes évoqués précédemment, cependant, les données sont rares et ces déclarations subjectives peuvent souffrir de biais de désirabilité.

Une autre possibilité est de construire des indices continus basés sur des dimensions de l'identité identifiée *a priori* ou non.

En outre, une mesure continue semble plus appropriée mais pose des questions quand à sa construction, à l'existence de données adéquates, à la reproductibilité des indicateurs pour des comparaisons dans le temps et l'espace.

Nous avons proposé une mesure continue du genre basée sur les pratiques culturelles des français. Notre indice converge vers les indices d'échelle dans la mesure où il indique que la

plupart des individus ne sont pas totalement conformes, que peu d'individus s'écartent des normes (entendues dans notre étude comme un comportement moyen), que ces divergences ou convergences sont reliées aux caractéristiques socio-économiques des individus.

Enfin, nous avons proposé d'étudier empiriquement la relation entre identité et satisfaction. Celle-ci n'apparaît pas comme évidente et est sensible au niveau de revenus des individus ainsi qu'à leur sexe biologique.

Les premiers résultats observés indiquent une relation négative entre distance aux normes de genre et satisfaction individuelle pour les femmes, cette relation n'est pas aussi prononcée chez les hommes.

La satisfaction (en termes de disposition de son temps libre) semble décroître avec l'augmentation de la distance aux normes chez les individus aux revenus les plus faibles.

Ces résultats sont préliminaires et méritent d'être approfondis.

References

- Akerlof, George A., and R. Kranton. 2000a. "Economics and Identity." *Quarterly Journal of Economics* 115: 715–53. <https://doi.org/10.1162/003355300554881>.
- Akerlof, George A., and Rachel E. Kranton. 2000b. "Economics and Identity." *The Quarterly Journal of Economics* 115 (3): 715–53. <https://www.jstor.org/stable/2586894>.
- Akerlof, George A., and Rachel E. Kranton. n.d. "Identity and Schooling: Some Lessons for the Economics of Education," 36.
- Austen-Smith, D., and Roland G. Fryer. 2005. "An Economic Analysis of 'Acting White'." *Quarterly Journal of Economics* 120: 551–83. <https://doi.org/10.1093/QJE/120.2.551>.
- Bem, Sandra L. 1974. "The Measurement of Psychological Androgyny." *Journal of Consulting and Clinical Psychology* 42 (2): 155–62. <https://doi.org/10.1037/h0036215>.
- Bénabou, R., and J. Tirole. 2004. "Incentives and Prosocial Behavior." *IZA Institute of Labor Economics Discussion Paper Series* null: null. <https://doi.org/10.2139/ssrn.639043>.
- Chakravarty, Surajeet, Miguel A. Fonseca, S. Ghosh, and S. Marjit. 2015. "Religious Fragmentation , Social Identity and Cooperation : Evidence from an Artefactual Field Experiment in India *." <https://doi.org/10.1016/j.euroecorev.2015.12.006>.
- Charness, G., Ramón Cobo-Reyes, and Natalia Jiménez. 2014. "Identities, Selection, and Contributions in a Public-Goods Game." *Games Econ. Behav.* 87: 322–38. <https://doi.org/10.1016/J.GEB.2014.05.002>.
- Chen, Yan, and S. Li. 2009. "Group Identity and Social Preferences." *The American Economic Review* 99: 431–57. <https://doi.org/10.1257/AER.99.1.431>.
- Cipriani, Enzo, Charles-Edouard Giguère, Eugénie Samson, Ioana Cotocea, and Robert-Paul Juster. n.d. "Comment mesurer indirectement le genre en recherche sur les humains?" 21.
- Fryer, Roland G., Roland G. Fryer, and Paul Torelli. 2010. "An Empirical Analysis of 'Acting White'." *Journal of Public Economics* 94: 380–96. <https://doi.org/10.1016/J.JPUBECO.2009.10.011>.
- Magliozzi, Devon, Aliya Saperstein, and Laurel Westbrook. 2016. "Scaling up: Representing Gender Diversity in Survey Research." *Socius* 2 (January): 2378023116664352. <https://doi.org/10.1177/2378023116664352>.
- Octobre, Sylvie. 2008. "Loisirs culturels et construction du genre au sein de la famille:" *Agora débats/jeunesses* N° 47 (1): 98–110. <https://doi.org/10.3917/agora.047.0098>.
- Pelletier, Roxanne, Blaine Ditto, and Louise Pilote. 2015. "A Composite Measure of Gender and Its Association with Risk Factors in Patients with Premature Acute Coronary Syndrome." *Psychosomatic Medicine* 77 (5): 517–26. <https://doi.org/10.1097/PSY.000000000000186>.

- Shih, Margaret, Todd L. Pittinsky, and Amy Trahan. 2006. "Domain-Specific Effects of Stereotypes on Performance." *Self and Identity* 5 (1): 1–14. <https://doi.org/10.1080/15298860500338534>.
- Trachman, Mathieu. 2022a. "Très masculin, pas très féminine. Les variations sociales du genre:" *Population & Sociétés* N° 605 (10): 1–4. <https://doi.org/10.3917/popsoc.605.0001>.
- . 2022b. "Très masculin, pas très féminine. Les variations sociales du genre:" *Population & Sociétés* N° 605 (10): 14. <https://doi.org/10.3917/popsoc.605.0001>.