Chapitre 4

Results

4.1 Descriptive Statistics

In this experiment 17 participants were enrolled to participate in the 12 belief formation tasks for a total of 204 observations. The experiment was run using the experimental software z-Tree and a session lasted about 90 minutes. Subjects were paid the same amount of 25 CAN. Table 4.1 presents descriptive statistics of participants. The mean age is 30 years and about 41% of

TABLE 4.1 – Descriptive statistics

Variables	Mean	Std.Dev	Min	Max
Age	30	6	22	47
Gender				
Male	0.59	0.49	0	1
Female	0.41	0.49	0	1

subjects are women. The subjects are either undergraduate or graduate. By pooling the data across treatments, table 4.2 presents summary statistics for all treatments and reveals that subjects evaluation of the correlation by presentation differ to the true value of correlation in some treatments. Thus, table 4.2 presents a sufficient amount of correlation neglect between individuals and treatments. We focus our analysis on correlation neglect at individual-level in the next section.

4.2 Heterogeneity in correlation neglect

Because table 4.2 reveals that some subjects neglect correlation, we develop a measure of individual correlation neglect in order to investigate this heterogeneity. Our design allows us to estimate individual's correlation neglect parameter by informational presentation k and treatment j, $\chi_i^k = med(\phi^{j,k} - \phi_i^{j,k})$ for each level of correlation. At the individual level,

TABLE 4.2 – Median correlation by treatment and informational presentation

True correlation	Median Subjective Correlation		
value	structural	Empirical without order	Empirical with order
0.33	0.00	0.15	0.20
0.20	0.20	0.18	0.20
0.11	0.12	0.00	0.00
0.005	0.00	0.00	0.00

The table presents a summary statistics of subjective correlations across treatments for each presentation.

graphics 6-8 reveal that neglecting the correlation somewhat depends on the presentation of information [1].

Figure 1 provides kernel density estimates of the distribution of these median correlation neglect parameters by informational presentation. The plots reveal 2 spikes for structural presentation ; one around zero for the vast majority of subjects (they behaves approximately rational) and one around 0.2 (correlation neglecters). For the others types of informational presentation, we observe 3 spikes when the majority of subjects also behave rational with the spike around 0. The others spike suggest the presence of different types of individuals.

This procedure however ignores the variability in subjects tendency to neglect correlation. Figure 1 suggests the existence of different types of individuals who neglect the correlation at different level of informational presentation. For each type of presentation, some participants behave like if they completely ignore the correlation between variables. For the purpose of finite mixture model, we suppose that each participant is characterized by a set of two-dimensional types (χ_t^k, σ_t^k) with $t \in \{1, ..., T\}$, and $k \in \{1, 2, 3\}$, where the population weights π_t are estimated along with (χ_t, σ_t). σ_t^k is the variance of individual type t in informational presentation k. The correlation neglect parameter of subject i in round j for presentation k can be expressed as $\chi_i^{j,k} = \chi_t^k + \mu_i^{j,k}$, $\mu_i^{j,k} \sim N(0, \sigma_t^k)$. The likelihood contribution of individual i in presentation k is given by :

$$L_i(\chi^k, \sigma^k, \pi^k) = \sum_{t=1}^{T} \pi_t \prod_{j=1}^{17} P[\chi_i^{j,k} = \chi_t^k + \mu_i^{j,k} | \chi_t^k, \sigma_t^k]$$

The grand likelihood is obtained by summing the logs of the individual likelihood contributions. The model generates different results depending on the number of types T included. The estimations are ran for up to $T = 5$ and we report results for up to $T = 3$ for the two different historical data presentations (with and without order) and for up to $T = 4$ for the structural presentation of the information. For the others types, the results are significantly similar for the case $T = 3$ and $T = 4$ but the likelihoods are scarcely improved relatively.

1. We investigate this issue in section 4.4.1

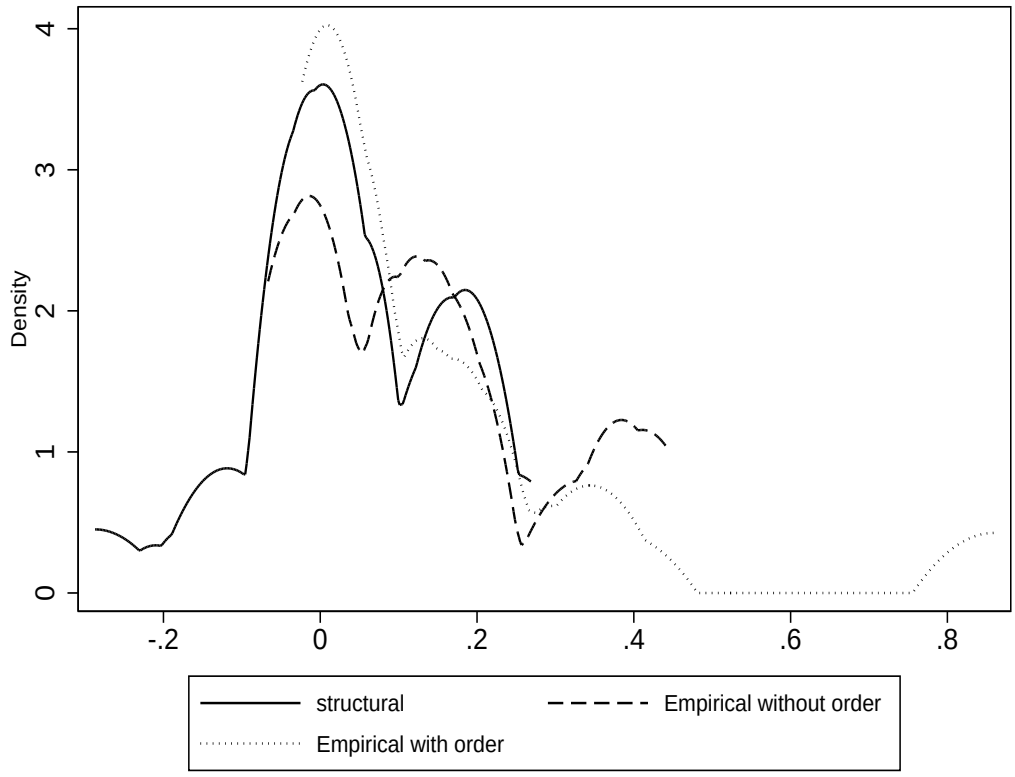

FIGURE 4.1 – Heterogeneity in Correlation Neglect

The estimates show that when we allow for one type of individuals in the experiment and for each informational presentation, the variance estimated is high. This mean that, the model with one class of individuals masks a considerable degree of heterogeneity. By allowing the existence of two types of individuals, the model fit increases but the variance is still higher in some groups. The model suggests that the data can be explained as a mixture of two differents groups of subjects. For one group, the estimates generate a correlation neglect parameter close to 0 (rational subjects). The second group is characterized by a large amount of correlation neglect. The high variances in some groups predict the presence of further sub-populations in the data. We allow for three types of individuals in the experiment. If we allow for more than three types in our historical presentations and more than four types in structural presentation the models fits increase but not dramatically and the parameters estimated for some groups remain unchanged. This individual-level analysis shows that subjects tendency to ignore correlation masks a considerable heterogeneity and the results are similar to the inference from Figure 4.1.

TABLE 4.3 – Structural

		Parameter			Goodness of fit		
Model	Type	χ	σ	π (%)	LL	AIC	BIC
T=1	t=1	0.04	0.14	100	40.12	-78.25	-76
		(0.02)	(0.02)				
	t=1	-0.28	0	6			
T=2		(0)	(0)	(0.03)	724.2391	-1442.478	-1435.82
	t=2	0.06	0.11	94			
		(0.014)	(0.01)	(0.03)			
	t=1	-0.28	0	6			
		(0)	(0)	(0.03)			
T=3	t=2	-0.02	0.05	59	735.9765	-1459.953	-1446.636
		(0.008)	(0.006)	(0.06)			
	t=3	0.18	0.044	35			
		(0.01)	(0.0065)	(0.06)			
	t=1	-0.28	0	6			
		(0)	(0)	(0.03)			
	t=2	0.004	0.002	47			
T=4		(0.0004)	(0.0003)	(0.06)	824.5472	-1631.094	-1611.119
	t=3	0.18	0.04	35			
		(0.009)	(0.006)	(0.06)			
	t=4	-0.12	0.013	12			
		(0.005)	(0.034)	(0.04)			

17 subjects, standard errors in parentheses.

TABLE 4.4 – Empirical without order

		Parameter			Goodness of fit		
Model	Type	χ	σ	π (%)	LL	AIC	BIC
T=1	t=1	0.13	0.16	100	26.600	-51.198	-48.978
		(0.02)	(0.03)				
	t=1	0.05	0.09	77			
T=2		(0.013)	(0.01)	(0.05)	39.523	-69.046	-57.948
	t=2	0.38	0.05	23			
		(0.014)	(0.01)	(0.05)			
	t=1	-0.014	0.044	47			
		(0.008)	(0.005)	(0.06)			
T=3	t=2	0.38	0.055	23	91.186	-166.372	-148.616
		(0.014)	(0.05)	(0.05)			
	t=3	0.15	0.003	30			
		(0.0008)	(0.0005)	(0.055)			

17 subjects, standard errors in parentheses.

TABLE 4.5 – Empirical with order

		Parameter			Goodness of fit		
Model	Type	χ	σ	π (%)	LL	AIC	BIC
T=1	t=1	0.13	0.21	100	8.073	-14.14576	-11.926
		(0.03)	(0.05)				
	t=1	0.05	0.07	75.7			
T=2		(0.012)	(0.009)	(0.075)	36.106	-62.213	-51.116
	t=2	0.4	0.27	24.3			
		(0.09)	(0.05)	(0.075)			
	t=1	0.055	0.07	82.5			
		(0.01)	(0.0075)	(0.05)			
T=3	t=2	0.86	0	6	185.695	-359.390	-346.073
		(0.00)	(0.00)	(0.03)			
	t=3	0.34	0.033	11.5			
		(0.011)	(0.008)	(0.04)			

17 subjects, standard errors in parentheses.

4.3 Cognitive effort and learning over time

This section investigates the relationship between correlation neglect and subjects' response times commonly used as proxy of cognitive effort (Rubinstein (2007)). Because of cognitive costs, subjects might develop a solution strategy and opt for simplifying heuristic. Table 4.6 provides the results of panel regression with random effects and heteroskedasticity-robust standard errors of correlation neglect parameter for each treatment on subjects' response time. Then, we check if subjects learn about correlation over time. The results show that correlation neglect is not associated with response time. A longer time spent on a task doesn't affect subject's tendency to neglect correlation.

TABLE 4.6 – Cognitive effort and learning over time

Dependent variable : Correlation Neglect Parameter	
Time Treatment	-0.00002
	(0.00003)
Time Trend	-0.002
	(0.009)
Time Treatment# Time Trend	0.00001
	(0.00002)
1 if Female	0.18
	(0.13)
Age	0.093**
	(0.038)
Age^2	-0.0015***
	(0.0006)
Education	-0.014
	(0.045)
Constant	-1.36**
	(0.63)

Controls variables include age, gender, level of education. *p < 0.10, **p < 0.05, ***p < 0.01.

Do subjects learn about correlation over time ? the results suggest that correlation neglect doesn't become smaller over time.

4.4 Mechanisms underlying correlation neglect

4.4.1 Correlation neglect and informational presentation

As shown in table 4.7 and graphics A.1-A.4 (appendix A), correlation neglect is associated with informational presentation. Some presentations allow to better recognize correlation between states variables than others. To investigate this issue, we regress individual correlation neglect parameter on informational presentation.

TABLE 4.7 – Correlation neglect and informational presentation

	Median correlation neglect	
	(1)	(2)
2 if empirical without order	0.14***	0.065*
	(0.04)	(0.04)
3 if matrix form	0.04	0.001
	(0.04)	(0.04)
Constant	0.0025	0.05
	(0.03)	(0.17)
controls	No	Yes
Obs.	204	204
Pseudo R^2	0.05	0.08

Median regression, standard errors in parentheses. Controls variables include age, gender, level of education. *p < 0.10, **p < 0.05, ***p < 0.01.

The results show that relative to the baseline presentation (structural presentation), observing historical data without ordering significantly increases subjects tendency to neglect the correlation. There is no significant difference between structural presentation and historical data presentation with ordering possibility. This results reveal that structural and empirical with order presentations are better to reduce individuals' correlation neglect.

4.4.2 Number of historical informations

Our endogenous treatment allows subjects to decide the number of past observations they observe. A good strategy is to observe all the data up to 100 draws and report the same distribution in his beliefs formation tasks. Tables 4.8-4.11 give the number of subjects who simultaneously choose a certain number of draws in the two empirical treatments (with and without ordering) for each level of true correlation. The tables reveal that the majority of individuals (in bold) observe up to 100 draws from the 2 informational presentations. However, we are not able to tell if the same subject choose the same number in all empirical treatment during the experiment. We investigate this issue by regressing the number of observations observed by each subject on time trend. The results in table 4.12 suggest that the number of observed informations doesn't increase significantly over time. Some of subjects keep the same number of observations during experiment. The results also suggest that level of education and gender affect the number of informations. Most educated individuals tend to ask for more informations than less educated one and women ask for less informations than men.

TABLE 4.8 – correlation=0.33

		Emp. With order			
		10	20	100	Total
Emp. Without order	10	2	1	0	3
	20	3	0	1	4
	30	1	0	2	3
	100	0	0	**7**	7
	Total	6	1	10	17

TABLE 4.9 – correlation=0.20

		Emp. With order				
		10	20	40	100	Total
Emp. Without order	10	1	1	0	0	2
	20	2	1	1	2	6
	30	0	0	0	1	1
	60	1	0	0	1	2
	100	1	0	0	**5**	6
	Total	5	2	1	9	17

TABLE 4.10 – Correlation=0.11

		Emp. With order				
		10	20	30	100	Total
Emp. Without order	10	3	1	1	1	6
	20	1	2	0	1	4
	30	0	0	0	1	1
	100	0	0	0	**6**	6
	Total	4	3	1	9	17

We are also interest to investigate the effect of the number of historical observations allow by our two empirical presentations on subjects tendency to neglect the correlation. In table 4.13, we run a median regression of correlation neglect parameter on the number of informations. The table reveals that more information tend to reduce correlation neglect. So, it is important to observe the maximum of information available before forming his beliefs.

TABLE 4.11 – Correlation=0.005

		Emp. With order			
		10	20	100	Total
Emp. With order	10	1	0	0	1
	20	1	2	2	5
	30	1	0	0	1
	40	0	0	1	1
	90	0	0	1	1
	100	3	1	**4**	8
	Total	6	3	8	17

TABLE 4.12 – Number of historical information over time

	Dependent variable : Number of histirical informations					
	Empirical without order		Empirical with order		Emp. without order + Emp. with order	
	(1)	(2)	(3)	(4)	(5)	(6)
Time	0.66	0.57	0.07	-0.02	0.00	0.31
	(1.23)	(1.2)	(0.97)	(0.0001)	(0.45)	(0.74)
Age		-0.88		-2.4***		-1.65***
		(0.83)		(0.8)		(0.57)
1 if female		-18.2		-17.75*		-17.8**
		(11.13)		(10.6)		(7.65)
Educational level		2.23		9.22*		5.8
		(5.62)		(5.35)		(3.86)
constant	48.70***	80***	60***	127***	52.9***	103.4***
	(10)	(28)	(11.5)	(26.7)	(0.04)	(19.22)
Obs.	68	68	68	68	136	136
R^2	0.004	0.11	0.0001	0.30	0.0023	0.19

Median regression, standard errors in parentheses. Controls variables include age, gender, level of education. *p < 0.10, **p < 0.05, ***p < 0.01.

TABLE 4.13 – Correlation neglect and number of past observations

	Dependent variable : Correlation Neglect Parameter			
	Empirical without order		Empirical with order	
	(1)	(2)	(3)	(4)
Numb. Past observations	-0.002	-0.002*	-0.002**	-0.0008
	(0.0008)	(0.0007)	(0.0006)	(0.0005)
Constant	0.14**	0.56***	0.17***	0.06
	(0.06)	(0.16)	(0.05)	(0.13)
controls	No	Yes	No	Yes
Obs.	68	68	68	68
Pseudo R^2	0.03	0.11	0.07	0.08

Controls variables include age, gender, level of education. *p < 0.10, **p < 0.05, ***p < 0.01.

Chapitre 5

Applications

5.1 Mother's age at first childbearing and Child mortality

Young maternal age is associated with adverse birth outcomes for child and mother (Omar et al. (2010) ; Wang et al. (2012) ; Shrim et al. (2011) ; Kang et al. (2015) ; Fraser et al. (1995)). Measuring people *"ignorance"* of this correlation may help for news stragies to reduce adverse birth outcomes for children and mothers. Recent literature (Diarra and Dessy (2017)) on the determinants of the demand of child bride suggests that people ignores the correlation between women's age at first birth and the level of mother mortality, or fails to factor this information in their marriage decision. Our design allows us to measure individual correlation neglect by eliciting their expectation about the age of bride at first childbearing and the level of infant mortality [1].

We elicit these distributions using natural frequency questions and asking for probability for outcome intervals. People are asked to share their evaluation out of 100 randomly selected women at age 12 or less, how many will have their first child at different age groups. This question elicits their subjective distribution of age at first birth.

Then, people are told to share their expectation out of 100 randomly selected newborn-babies, how many will died before reaching their first birthday. This question assesses subjective distribution about infant mortality.

finally, we elicit beliefs about the joint distribution of two variables. Subjects have to answer these questions in the following way :

A.) *I think that out of the 100 randomly selected women at age 12 or less,*

(1) *"...d*1*... will have their first birth before* 18 *years old."*

(2) *"...d*0*... will have their first birth from* 18 *years old. "*

B.) *I think that out of the 100 randomly selected newborn-babies,*

1. Because mother mortality is not observed in the data in order to compute an empirical correlation with mother mortality

(1) *"...b1... will died before their first birthday,"*

(2) *"...b0... will be alive until their first birthday,"*

C.) *I think that out of the ...d1... women who will have their first birth before* 18 *years old,*

(1) *"...d1b1... will have their baby died before reaching age 1,"*

(2) *"...d1b0... will have their baby alive until their first birthday,"*

D.) *I think that out of the ...d0... women who will have their first birth from* 18 *years old,*

(1) *"...d0b1... will have their baby died before reaching age 1,"*

(2) *"...d0b0... will have their baby alive until their first birthday,"*

5.2 Financial Allocation Decisions

In order to measure people perception of the correlation in assets returns when making their financial allocation decisions, our measure using beliefs elicitation can be apply. For simplicity, like Kallir and Sonsino (2009), portfolio contains two assets A and B with two levels of returns : "*high*" and "*low*". Information on the joint distribution of returns is present in the form of empirical frequencies for 12 preceding periods as presented in table 5.1. Subjects are told to predict returns for 100 additional periods, under the assumption that future returns are sampled from the empirical distribution.

		Asset B	
		high	low
Asset A	high	5/12	1/12
	low	1/12	5/12

TABLE 5.1 – Joint distribution

The corresponding questions eliciting the joint distribution of 100 future returns for assets A and B are :

1. *"How many out of the 100 additional periods do you think that assets A and B will have simultaneously high returns levels ?"*
2. *"How many out of the 100 additional periods do you think that assets A and B will have simultaneousl low returns levels ?"*
3. *"How many out of the 100 additional periods do you think that asset A will have high return and low return for asset B ?"*
4. *"How many out of the 100 additional periods do you think that asset A will have low return and high return for asset B ?"*

By varying the correlation in assets returns and asking for the same questions as above, we can analyse subjects responses to change in correlation.

5.3 Pre-election Polls

Subjects are told to predict the results of an election. They have a choice between two options A and B. Each option has two levels of cost : high cost or low cost. The two options represent the two candidates for presidential race. There is 4 types of individuals in this population characterized by 4 different colors : blue, dark green, pale green and yellow. Each individual belongs to only one type of color. Subjects in the experiment observe the decisions of 100 individuals, randomly selected in the whole population (25 individuals per color). The results are presented in table 5.2. Subjects are told that this table represents the joint distribution between the choice of an option and his cost. Subjects have to predict to predict the choices of 100 additional people sampled in the same population such that 25 persons were drawn in each group. The following questions are used to assess their beliefs for 100 additional observations :

	Blue				Pale Green		
	high cost	low cost	Total		high cost	low cost	Total
Option A	2	20	22	Option A	12	3	15
Option B	1	2	3	Option B	5	5	10
Total	3	22	25	Total	17	8	25

	Dark Green				Yellow		
	high cost	low cost	Total		high cost	low cost	Total
Option A	1	11	12	Option A	4	1	5
Option B	1	12	13	Option B	18	2	20
Total	2	23	25	Total	22	3	25

TABLE 5.2 – subjects choices for each type of individual

1. *"How many out of the 100 additional people do you think, will choose option A and will pay high costs ?"*
2. *"How many out of the 100 additional people do you think, will choose option A and will pay low costs ?"*
3. *"How many out of the 100 additional people do you think, will choose option B and will pay high costs ?"*
4. *"How many out of the 100 additional people do you think, will choose option B and will pay low costs ?"*

Subjects who perceive the correlation between variables should fill in exactly by summing the number of individuals across groups who are in the same situation in the historical distribution that they observed. This design allow to measure, for each subject, his level of correlation neglect.

Conclusion

This paper has provided a sequence of belief formation tasks by varying the level of correlation between tasks and has demonstrated that people neglect correlation when they face some types of informational presentation. In this paper, we suppose correlation neglect to be an individual characteristic of a person and we suggest an empirical measure. Our measure is based on a sequence of laboratory experiments and tests for 3 types of informational presentation. First, subjects learn about the structure of decision problem and make their predictions ; second, they observe historical data of state variables without ordering possibility and third, they have a possibility to sort historical data.

Our results suggest a good amount of heterogeneity in correlation neglect at individual-level analysis. Two types of presentation allow to reduce subjects' tendency to neglect correlation : structural presentation and historical presentation with ordering. Women are likely to neglect correlation more than men and the level of education does not significantly affect their neglect.

Empirical data analysis suggests that individuals may observe all the information available in order to form their beliefs about state variables. Subjects don't learn about correlation over time and cognitive effort doesn't affect the correlation neglect parameter.

Our strategy, because we are able to measure correlation neglect at individual-level without making any ancillary hypothesis about individual's preferences and without observing their decisions making process can be use as a general measure of correlation neglect when asking people's expectations about the distribution of state variables.

Although we propose a simple measure of individual correlation neglect that can be applied in various domains, some authors suggest a measure that is specific to a domain (investment decision and portfolio choice problem, auctions market setting, etc.). These measures present the correlation either by the structure of information (correlation neglect in informational source or common source of information) or by observing historical data (e.g returns) and compute correlation neglect based on individual's investment decision.