Doctoral Thesis

UNIVERSITY OF TRENTO

DOCTORAL SCHOOL OF SOCIAL SCIENCES

DOCTORAL PROGRAMME IN ECONOMICS AND MANAGEMENT

An eye tracking exploration of cognitive reflection in consumer decision-making

A dissertation submitted to the Doctoral School of Social Sciences in partial fulfillment of the requirements for the Doctoral degree (Ph.D.) in Economics and Management.

Alessia

Dorigoni

April, 2019

Advisor: Professor Nicolao Bonini - University of Trento

To my mom and my brother, thank you. To my daughter, I love you. To my dad, I miss you.

Alessia

Acknowledgements

I thank my colleagues, especially Jonathan Nelson, Julie Sriken, Sebastien Payet and Michele Graffeo, who provided insight and expertise that greatly assisted the research; and a special thanks to Luca Polonio, Kenneth Holmqvist, Diederick Niehorster, Tim Cornelissen, Roy Hessels, Ignace Hooge and Ricardo Matos for assistance with the eye tracker and for comments that greatly improved the manuscripts.

I would like to thank Professor Maya Bar Hillel, Professor Ilana Ritov and Professor Umberto Martini for their valuable suggestions for improvement.

I would like to show my warm thanks to Jason Rajsic and Annalisa Cadonna. I have learned a lot from you and you have contributed to high quality research.

I wish also to present my special thanks to my supervisor, Dr. Bonini, for his vital support, assistance and for his helpful hints. His encouragement made it possible to learn important research tools and to achieve the goal of this thesis.

And last but not least, I would like to show my gratitude to my friends, in particular Michela Gandini, who assisted me by extending their support morally and emotionally; and to my family members, especially to my mom and my brother: I couldn't have gotten through the last three years without you.

Abstract

The works presented in this thesis are the result of the experiments conducted in the Cognitive and Experimental Economics Laboratory (CEEL) and in the Consumer Neuroscience Laboratory (NCLab) of the Economics and Management Department at the University of Trento. The aim of this research is to study the influence of cognitive impulsivity on commercial problem-solving and consumer decision-making. We focused on the attentional aspects related to the decision-making process as analyzed by the eye movements. The first section will present the main topic of the thesis, the key tool used to conduct the experiments (eye tracker) and the three papers; the latter will compose the second, third and fourth chapter. All the chapters have a common thread: to shed light on the cognitive aspects of problem-solving and their implications for the consumer decision-making process as analyzed through gaze behaviour.

- The aim of the first paper, "The role of numeracy, cognitive reflection and attentional patterns in commercial problem-solving" by Dorigoni, Polonio, Graffeo and Bonini, is to analyze the predictive power of two important cognitive abilities, numeracy and cognitive reflection, in two different problem solving scenarios with high numerical components.
- The aim of the second paper, "Getting the best deal: Effects of cognitive reflection on mental accounting of choice attributes" by Dorigoni, Cadonna and Bonini is to understand if people with low cognitive reflection are more prone to mental accounting across attributes of the same product; low cognitive reflectors do not integrate all the attribute costs and consequently they do not always choose the best deal.
- The aim of the third paper, "Cognitive reflection and gaze behaviour in visual tasks" by Dorigoni, Rajsic and Bonini is to demonstrate that cognitive reflection has predictive power on heuristics and biases related to perceptual and visual tasks. This result is extremely important because it reflects a different disposition to see and analyze the information depending on the cognitive impulsivity.

Keywords: *cognitive reflection, consumer decision-making, attentional aspects, heuristics and biases.*

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

CHAPTER 1

Introduction

Most people consume products and use services daily; most of the time the decisions that they have to make are consumption decisions and in many moments of our lives we act as consumers who make hundreds of decisions. This role of decision-maker is extremely crucial for our living standards so it is important to understand what allows us to fulfill this role. How we analyze information is an important part of the decision-making process. In this thesis, the focus is on the role of cognitive reflection on consumer decision-making because cognitive impulsivity leads to miserly information processing. We decided to collect eye movement data in order to better understand what drives the attention of the participants and how they analyzed and combined pieces of information.

In the Introduction I will present:

- 1. Cognitive reflection: a short literature review, focusing on the foundations of this construct and on its predictive power in different fields;
- 2. Eye Link 1000 Plus: an explanation of the fundamental principles of the tool used to record gaze behaviour;
- 3. Summary of Empirical Studies: a presentation of the aim of three papers that will be shown in the second, third and fourth chapter.

1. Cognitive reflection

1.1. What cognitive reflection is

The typical approach to introduce the notion of cognitive reflection is to use Frederick's initial definition ([1]). He conceived cognitive reflection as "a simple measure of one type of cognitive ability, the ability or disposition to reflect on a question and resist reporting the first response that comes to mind" (pp. 25).

Accordingly with this definition, on the one hand this cognitive ability allows for *resisting* to the first intuition on how to solve the problem; on the other hand for *reflecting* on the question of the problem in order to find a different solution. Cognitive reflection is the tendency to override an automatic, and incorrect, first answer to the problem, and to engage in further reflection that leads to the correct response. The roles of intuition and reason are explained by the dual process theories ([2], [3]) that consider cognitive reflection as the ability to stop the first intuition that comes automatically, with no effort (System 1) in order to move slowly in another direction, with control and effort (System 2) that could lead to the correct response; it measures the degree of control of System 2 on System 1 or, in other words, whether reason takes the reins of intuition.

Consequently, the Cognitive Reflection Test (CRT, [1]) was described as a measure of the ability to control a high accessible (incorrect) answer in favour of a less accessible (correct) answer, which requires further deliberation. Some authors have highlighted the critical issues related to CRT that try to measure both *resisting* and then *reflecting* (on a different answer) at the same time. As has been pointed out by Baron and colleagues [4] "CRT might be considered a test of reflection-impulsivity (RI) that is concerned only with the amount of thinking, not its direction" (pp.265). Indeed, the cognitive reflection measure can be seen also as an impulsivity measure. In particular, the CRT measures monitoring of System 1 intuitions such that heuristic (incorrect) responses will be detected if cognitive reflection is high enough ([5]). Indeed, in the literature, as we will see below, sometimes the answers to the CRT are divided in three main groups: intuitive answers, non-intuitive answers irrespective of their correctness and correct answers, in order to distinguish the *resisting* step to the second part of the problem-solving process, where people search for a solution. In order to understand if the intuitive answer emerges also for the people that answer correctly, Travers and colleagues ([6]) measured the mouse cursor movements of participants; they demonstrated that even people who answered correctly were initially drawn towards the incorrect, intuitive answer. This shows that engaging further effortful processing leads to blocking the heuristic responses.

In my opinion the dual-system dichotomy is not the only duality that emerges when we consider the cognitive reflection. We have at least other two main dualities: (a) problem/person and (b) stop/research.

(a) Problem/person dichotomy. CRT items could be considered as problems that need to be solved by a person. A problem, starting from the definition of Duncker, arises when a living creature has a goal but she does not know how to reach this goal ([7]). This happens also for the CRT items; it is crucial to analyze the nature of the problem in order to understand the cognitive reflection construct ([8]). Accordingly with this point, three are the main aspects of the CRT that I will analyze in comparison with another type of problems, the insight problems, in order to highlight and to bring out the importance of these aspects.

a. The specificity of all the cognitive reflection tests is that the problem in itself triggers an answer, an intuition. In well-known insight problem we have an opposite situation; classic insight problems do not trigger an attractive alternative response. The intuition, the *Aha!* experience, the *Eureka effect*, is what the people are looking for, it is difficult to reach and most of the time the intuition comes to the mind after an effortful research of the solution.
b. However, if in the first case the automatic intuition leads to the wrong way (the incorrect intuitive answer), in the second is the opposite; the effortful intuition is the correct answer (the correct not intuitive answer).
c. Furthermore, The *Aha!* experience is generally described as a very nice experience in which the participants have the total certainty about the accuracy of their solution ([9]). In the cognitive reflection test, however, the intuition appears also with a strong certainty that the solution is correct.

So, in the insight problems the intuition is effortful, after deliberate research, perceived as correct as well as it actually is; in the cognitive reflection problems, the intuition is immediate, effortless and perceived as correct even if at the end it is not. In the insight problems the participants are mixed up trying to find the solution looking for a new perspective from which they should reconsider the problem; but this happens only when the participant realizes that the intuitive answer is wrong. In summary CRT seems at first glance to be similar to the insight problems but they differ in two main aspects:

1. in terms of the presence of an intuition before or after a deliberative thinking;
2. in terms of accuracy of this intuition.

In the first case an intuition occurs after strong research with a high cognitive load but at the end the intuitive conclusion is correct; and in the second case the alternative response is initially primed and then must be overridden because it is wrong. In both cases, however, the intuition is perceived as correct.

(b) Stop/research dichotomy. As already mentioned, the CRT measures not only the ability to stop, to block the first automatic intuition, but also to research in another direction the correct solution. About the "stop phase" there is a crucial point that concerns both, CRT and insight problem: the concept of functional fixedness. This term, which originated in Gestalt Psychology, was initially used by Duncker to define a mental block, a cognitive difficulty in using an object with a fixed function by experience, in a different and unusual way in order to find the solution to a problem. The intuition comes when there is a reorganization of the elements and their meaning in the problem-solving space that overcomes the functional fixedness. "The decisive points in thought-processes, the moments of sudden comprehension, of the *"Aha!"* of the new, are always at the same time moments in which such a sudden restructuring of the thought-material takes place, in which something 'tips over'" ([7] pp. 375). The fixedness appears related to two different elements in the two kinds of problems:

1. in the insight problems the fixedness is related to the function of an element (e.g. to remain in the box, in the 9 dots problem);
2. in the cognitive reflection test, the fixedness is linked with the anchor to a particular thought and intuition (we could call it *fixedness on thought;* e.g. in the CRT anchoring on the answer: 1$ the bat and 10 cent the ball).

The crucial part of the CRT is not that the decision is based on the first intuition but that the problem solver appears to be blocked, anchored, fixed on an initial intuition. The intuition in the CRT is an auto-generated limitation that restricts the pool of solutions that the person is considering and only if the problem-solver breaks this block she could consider other possible solutions. People with lower cognitive reflection anchor on the first information they consider, thus, being more susceptible to reference points as compared to people with higher cognitive reflection. Fixedness is negatively related to divergent thinking; people with higher cognitive reflection are more likely to exhibit higher remote associations, originality scores and lower fluency scores in divergent thinking ([10], [11], [12]) and actively open mind thinking (AOT, [4]).
If the stop moment is represented by the fixedness on the first intuition, the research moment is represented by the disposition to reflect, quoting Frederick, to search the solution in another direction. The first point is what better emerges from the definition of Frederick and the cognitive reflection construct hinges on it. The second point was less debated, taking for granted that once a person is able to stop her first, wrong, intuition then it is almost automatic that she will find the right way in order to reach the goal, the solution or in other words, to climb the hill ([13]). Other authors besides Baron and colleagues [4] tried to manage this problem; Campitelli and Gerrans ([14]) appointed three kinds of answers to the CRT: the counter-intuitive answers (divided into correct answer and incorrect answers except for the intuitive answer) in order to measure inhibition ability and an intuitive answer that represents the incapacity to inhibit. Related to this point, on the other side of the problem, there is the individual point of view. As Toplak and collegues reported in their paper ([8] pp. 1276), "the potency of the CRT as a predictor may derive from the fact that it taps both a cognitive ability dimension and a thinking disposition dimension". So the CRT does not only measure the cognitive ability but also a thinking disposition. This last point is extremely interesting because it is difficult to disentangle these two concepts using a unique measure. It is important to understand what matters more in this

test, which subset of cognitive abilities is related to this construct and how thinking disposition, as tendency to engage in fully disjunctive reasoning and tendency to seek alternative solutions and hypothesis, contributes. Anyway, CRT is able to measure the depth of processing and consequently the tendency towards miserly processing (important aspects in everyday life) in a more accurate way compared to other cognitive ability tests.
In summary I reported all the above-mentioned similarities and differences between CRT and insight problems (Table 1) and a representation of the dichotomies that emerge from the CRT (Image 1).

	Intuition	Correctness of the intuition	Perception of correctness	Fixedness
CRT	Automatic intuition	Incorrect	Perceived as correct	Fixedness on thought
Insight problem	Effortful intuition	Correct	Perceived as correct	Functional fixedness

Table 1. Similarities and differences between CRT and insight problems.

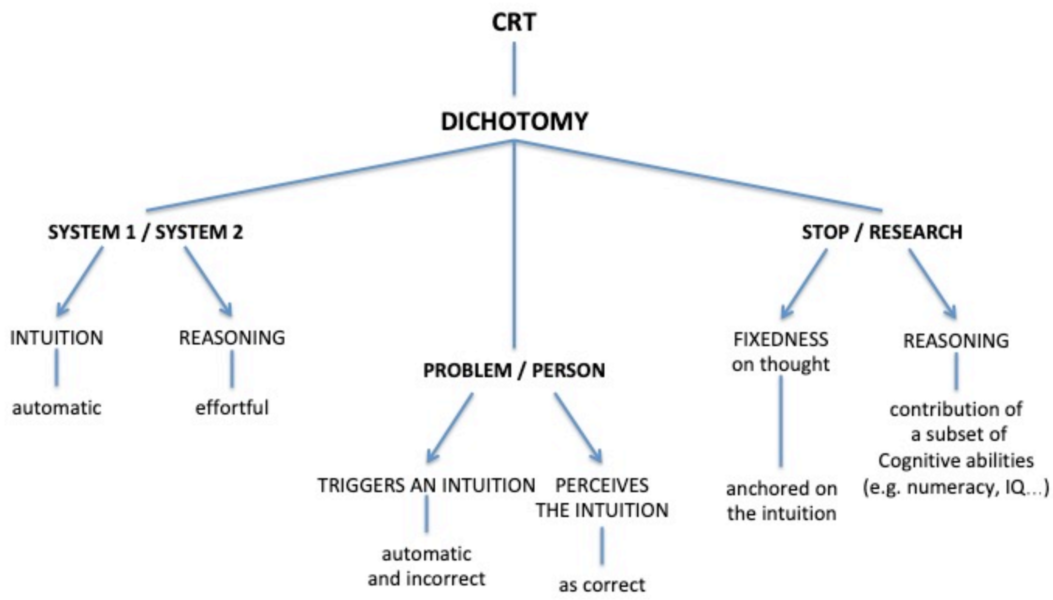

Figure 1. Representation of the dichotomies that emerge from the CRT.

Now I will analyze the different CRTs in the literature and the predictive power of the CRT in different situations.

1.2. Cognitive Reflection Tests

Analyzing Scopus source, the original paper by Frederick ([1]) was cited by 1254 articles from 2006 to 2018. As we can see from the Figure 2, the citations for this article increased during this period (from 1 citation in 2006 to 230 citations in 2018).

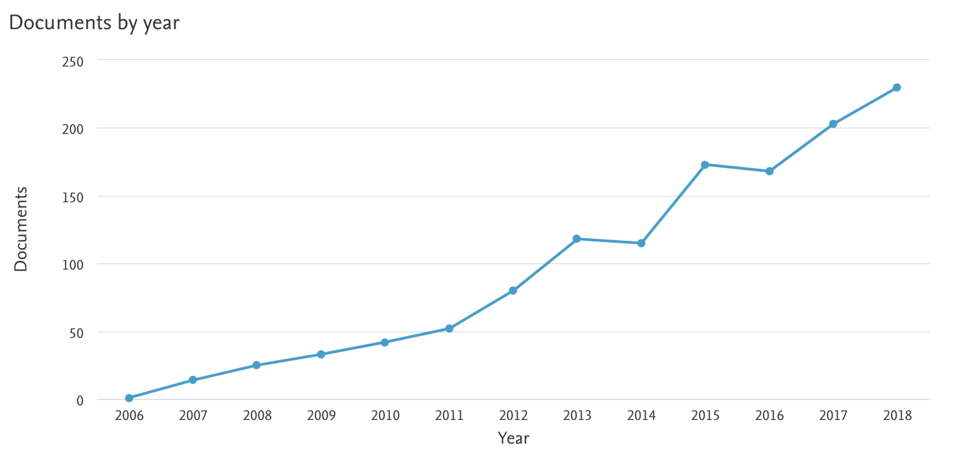

Figure 2. Documents by year that cited the original paper by Frederick about cognitive reflection.

An ever-increasing number of authors decided to study cognitive reflection and its impact in different fields (Figure 3).

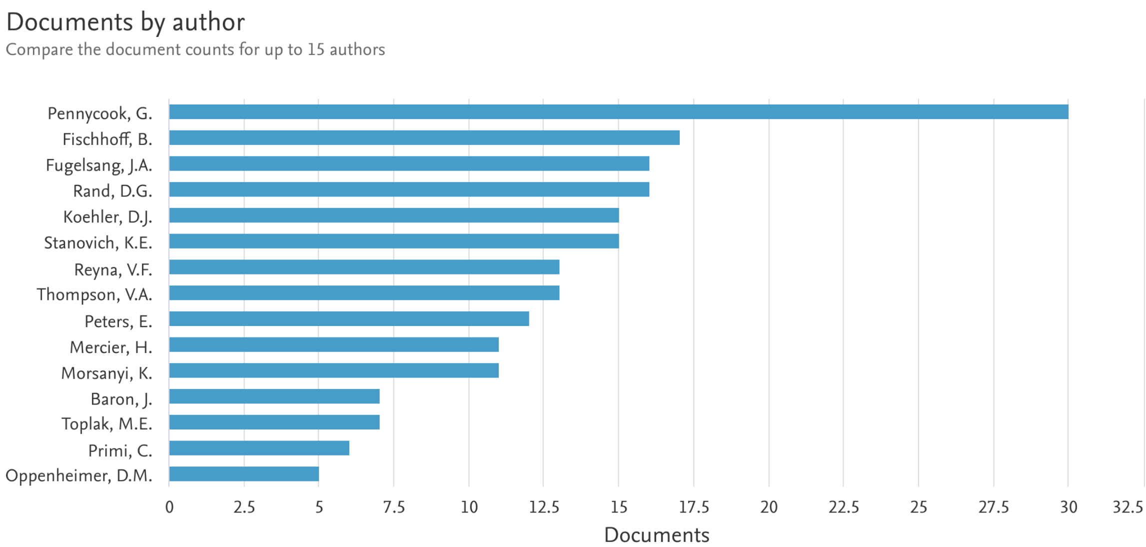

Figure 3. Documents by author that cited the original paper by Frederick about cognitive reflection.

In the literature, there are several ways to measure impulsivity but these are the main approaches:

(a) self-report. The most common is the Barratt Impulsiveness Scale (BIS-

11) [16]);

(b) performance based (Cognitive Reflection Test, [1]).

In this thesis, we considered and compare four performance based measurements of cognitive reflection (Table 2):

References	Test	Items
Frederick (2005)	CRT	A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost?
		If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
		In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?
Toplak et al. (2014)	CRT-4	If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together?
		Jerry received both the 15th highest and the15th the lowest mark in the class. How many students are in the class?
		A man buys a pig for $60, sells it for $70, buys it back for $80, and sells it finally for $90. How much has he made?
		Simon decided to invest $8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has: a. broken even in the stock market, b. is ahead of where he began, c. has lost money.
Primi et al. (2015)	CRT-L	If three elves can wrap three toys in hour, how many elves are needed to wrap six toys in 2 hours?
		Jerry received both the15th highest and the 15th lowest mark in the class. How many students are there in the class?
		In an athletics team, tall members are three times more likely to win a medal than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes?
Thomson and Oppenheimer (2016)	CRT-2	If you are running a race and you pass the person in second place are you in?
		A farmer had 15 sheep and all but 8 died. How many are left?
		Emily's father has three daughters. The first two are named April and May. What is the third daughter's name?
		How many cubic feet of dirt are there in a hole that is 3' deep x 3' wide x 3' long?

Table 2. Summary table of the CRTs used in this thesis.

1. The original CRT by Frederick [1] was a three-item test. Given its increasing popularity (it is becoming common knowledge among the university student population), many other tests were developed as alternatives. We analyze three CRTs and indirectly also the original one by Frederick; indirectly because, as we will see below, some of these cognitive reflection tests added the items to the original test.
2. Toplak and colleagues ([15]) added four items (CRT4) to the original three-item test (CRT3) with the result being a seven-item test (CRT7). The following aspects highlight the relationship between the CRT and the CRT4:
    a. the four new items had a quite high correlation with the classic three

(0.58) and, if we consider the seven-item scale (CRT7), it displayed substantial internal consistency;

b. considering the new four questions, the intuitive responses for them appear to be less dominant if we analyze the percentage of the 160 participants that gave the intuitive answers (the intuitive response in the original CRT was given 85.6%, 75.2%, and 60.0% of the time for the bat/ball, widgets, and lily pad problems, respectively; in the CRT4 the intuitive response was given 31.3%, 51.9%, 41.9%, and 53.1% of the time for the barrel, marks, pig, and stocks problems, respectively.

a. the CRT4 is simpler than the CRT3 (the probability to have a correct response in the first test is .24 and in the second is .17).

Toplak and colleagues examined also the items' ability to predict performance on:

a. seven rational thinking tasks (Belief Bias in syllogistic reasoning, Selection task, Denominator Neglect, Temporal Discounting, Otherside Thinking, Framing, Bias Blind Spot). The CRT4 is less predictive than the CRT in predicting rational thinking tasks;
b. cognitive ability (Wechsler abbreviated scales of intelligence);
c. four thinking dispositions (Need for Cognition, Actively Open minded Thinking, Superstitious Thinking, and Consideration of Future Consequences). The CRT4 is more predictive than the CRT in terms of thinking disposition.
d. temporal discounting; neither of them (CRT and CRT4) have predictive power of it.

One of the main findings of this study is that CRT4 does, in some cases, contribute incrementally to the predictive power of CRT7. One of the reasons for this result is that, as we said above, CRT4 has a higher correlation with thinking dispositions than the CRT3 does; the first is able to capture a dimension of cognitive reflection, indeed thinking disposition, and the second captures the cognitive ability part. CRT has its largest correlations with belief bias and denominator neglect. In these two tasks, the role of disjunctive reasoning and tendency to seek alternative hypothesis is crucial. Perhaps the CRT is not only a measure of the tendency towards the class of reasoning error that derives from miserly processing but also a measure of the disposition to search, the ability to find relation among elements and to represent and think about alternatives.

This new version of the test, as described in Primi et al. ([17]) has good reliability but four main problems:

2.1 one of the four new items has three responses and the subject has to choose one of them, with the possibility of a subject merely generating a response by chance;

2.2 for one item 31% of responders gave the heuristic response, whereas 46% of responders generated a different incorrect response (compared to the original CRT where the 60% of the respondents generated the heuristic response);

2.3 the dimensionality of the scale was not analyzed;

2.4 it is not clear how the authors identified which items were to be included in the scale.

The third CRT that we used and analyzed is the following:

3. Primi and colleagues ([17]) investigated not only the psychometric properties of the original CRT but also a new version of it with three new items (along with five items in the 2014 version, [18]), which they called CRT long (CRT-L). They defined the difficulty and discrimination parameters of the CRT items using the item response theory model and the authors investigated the validity of both the original CRT, and the CRT-L, by measuring their correlations with:
    a. intelligence (Advanced Progressive Matrices as a measure of fluid intelligence);
    b. mathematical and probabilistic reasoning, (probabilistic reasoning, numeracy, math fluency and subjective numeracy),
    c. reasoning ability (conditional probability and transitive inference);
    d. decision-making (Risk Seeking Behaviour, Intertemporal Behaviour, framing);
    e. thinking dispositions (Superstitious Thinking).

What they found about the original CRT is its unidimensionality, high discriminative power and difficulty. They also tested the unidimensionality of the CRT-L; concerning difficulty, the three new items are located one around, one below and one above the mean with a good level of discrimination. The new version of CRT appears to be similar to the original CRT in many aspects and components. Primi and colleagues make an open question of the necessity to develop a measure of cognitive reflection that shows weaker correlations with measures of numeracy or intelligence. Related to this point, we decided to include also the following test that tried to manage this problem.

4. The problem of the CRT being confounded with numeracy is an important point highlighted also in Thomson and Oppenheimer ([19]); they developed a four-item test (CRT-2) in order to increase the pool of available questions and to address numeracy confounding. Many studies have found a correlation between numeracy and the CRT suggesting that the CRT is a measure of both cognitive refection and numeracy. Some authors asked if we can define cognitive reflection as the propensity to give non-intuitive responses and numeracy as the ability to calculate the correct answer given that the intuitive answer was rejected ([5]). By definition, if we block our first intuition it is more likely that we would explore at least one more way to solve the problem, compared to those who are not able to inhibit their first intuition (stop/research dichotomy); but does the numeracy matter (and how far) to find the solution? In order to disentangle the roles of cognitive reflection and numeracy during the problem-solving process, Thomson and Oppenheimer used problems with less math components, "verbal CRT-type questions", that do not require a high degree of mathematical sophistication. They tested the new measure and its correlation with belief bias, numeracy scale, time preference, risk preference and need for cognition. The authors suggested that the belief bias items are similar to the CRT questions because in both tasks is required to override an intuitive tendency; both these tasks might measure the same construct ([19]). The result of this paper, in addition to the important role of belief bias that can also be a source of CRT items, showed that CRT-2 does display a significant relationship with numeracy, even if the items do not have a prevalent mathematical component, but with a really low dependence compared with the original CRT.

1.3. Cognitive reflection and numeracy

As we mentioned above, there is widespread study of the significant math components of the CRT ([14], [20]) and the nature of their relations ([5], [21]) in order to understand what cognitive reflection intrinsically is and whether it exists as an independent construct. Thus, in our studies we decided to analyze also the numeracy skill in order to understand its relationship with CRT in the consumer decision-making process.

The ability to comprehend and manipulate probabilistic and other numeric information is an important skill that allows for an understanding and correct usage of numerical information presented in text, tables, or charts. This construct, extensively studied also by OECD is defined by PIAAC as the ability "to access, use, interpret and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life [...] the inclusion of "engage" in the definition signals that not only cognitive skills but also dispositional elements, i.e. beliefs and attitudes, are necessary for effective and active coping with situations involving numeracy" ([22], pp. 34). This construct appears to be a simple measure but within itself it has many aspects that were widely studied in the literature.

As in the CRT we have two main approaches to measure this construct:

a) self-assessment scale to arithmetic skills. Fagerlin and colleagues ([23]) developed the subjective numeracy 6-point Likert scale, a subjective measure with four items measuring people's beliefs about their skill in performing various mathematical operations, and four items measuring people's preferences regarding the presentation of numerical information;
b) performance based; math fluency, captured by the numeracy scale, describes the ability to solve calculation and operation problems quickly [24]. The numeracy scale broadly studied in the literature and developed by Lipkus and colleagues in the well-known 11 items test ([25]) is an example of this kind of measurements. They expanded the Schwartz

scale ([26]) adding eight questions to it. The Schwartz scale was one of the first performance-based numeracy measures and it included three (one question assessing participants' understanding of chance and two questions asking the participants to convert a percentage to a proportion and vice versa).

In the second chapter (where the paper " The role of numeracy, cognitive reflection and attentional patterns in commercial problem-solving" is presented) we used the 8-item numeracy scale developed by Weller et al. ([27]). It is a shorter, psychometrically improved measure of numeracy. The scale assesses a broader range of difficulty and it is a better predictor of risk judgments compared with previous measures.

Recent reviews of the numeracy literature have found that compared with highly numerate individuals, those lower in numeracy are more likely to have difficulty judging risks and providing consistent assessments or utility, are worse at reading graphs, show larger framing effects, and they are more sensitive to the formatting of probability information; it was observed that being highly numerate can lead to increased effective reactions to numbers, or number comparisons, which, in turn, can result in optimal or sub-optimal decision-making ([25], [27]).

The link between CRT and mathematical skills is also supported by the literature ([28]) and it may be observed by a gender analysis. Male participants scored significantly higher than females on the CRT ([1], [29], [30], [31], [32], [33]); it has been argued that gender differences in the CRT might result from the mathematical component of the CRT, as gender effects might be reduced or eliminated when controlling for numeracy ([17]). Moreover, higher math anxiety (more frequent in female population) leads to worse performance in the cognitive reflection test ([34]). Even if it has been demonstrated that testosterone reduces cognitive reflection ([35]), further studied should be provided in order to understand the role of gender in the CRT performance.

In the second chapter we focused on numeracy because we have been asking an important question about the importance of cognitive reflection and numeracy: "In consumer problems involving high numerical components is numeracy vulnerability the only factor to play an important role?". We focused our attention on cognitive reflection because, considering the literature, numeracy is not the only ability that matters in these kinds of problems and people with high numeracy skills, who handle numbers daily, are affected by some biases and heuristics even in problems involving high numerical components ([36], [37], [38], [39]). So our idea of "overlapping vulnerabilities" was developed because we think that numeracy is not the only ability that plays a role in these kinds of problems; cognitive reflection has an important role too, because it predicts the control of the System 2 on the System 1.

What exactly the CRT measures is still subject to debate. The doubt that the test only measures mathematical abilities, and not cognitive reflection since the CRT contains three mathematically based problems, still persists. Further studies are needed at present in order to better understand this construct and to implement a test without any mathematical components in order to address the correlation between numeracy and CRT. It is also debated if CRT has predictive power or not ([5]) and in many studies this ability seems to not play a role independent of numeric abilities in the decision.

1.4. What Cognitive Reflection predicts

An increasing number of papers that consider and use CRT was written by a huge numbers of authors (Figure 4) in different fields (Figure 5).

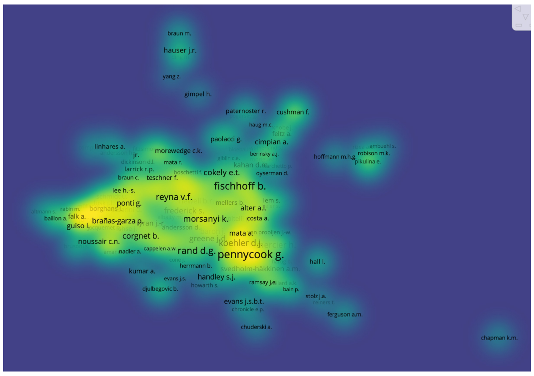

Figure 4. Density visualization of the citation analysis. The relatedness of items is determined based on the number of times they cite each other.

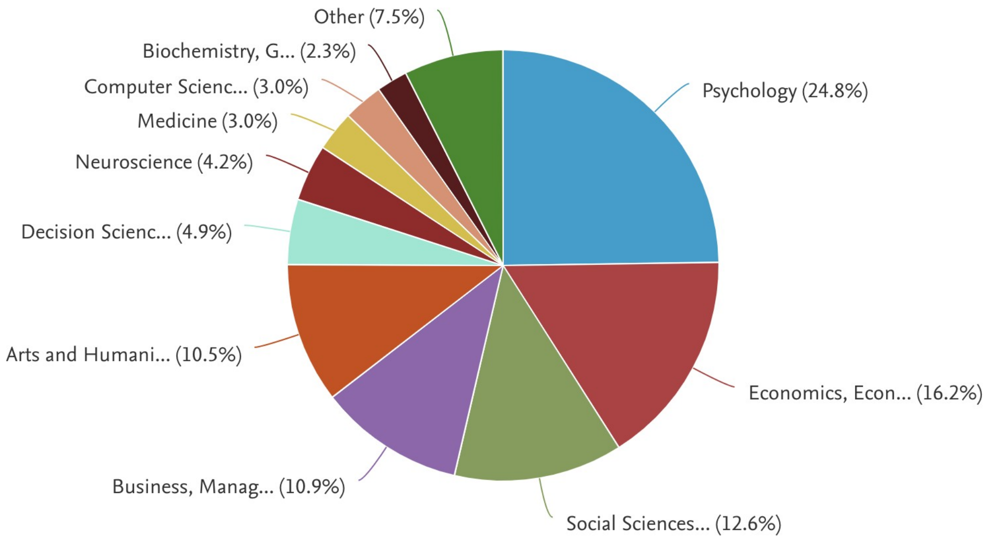

Figure 5. Documents by subject area.

In many studies, cognitive reflection was analyzed in combination with several other tasks that cover several fields. It is demonstrated to be a predictor of:

a) rational thinking and reasoning ability ([8], [15]);
b) decision-making skills, time and risk preferences (people with higher cognitive reflection is more likely to choose a future reward of greater value than a smaller immediate reward ([1]);
c) thinking dispositions ([40]) which in turn are related with economic behaviour and decision.
d) a much wider range of the heuristics-and-biases tasks and judgment-and-decision-making tasks ([8]);
e) conjunction fallacy, illusion of control, overconfidence, base rate fallacy, conservatism ([41]) and ratio bias ([42]).

Cognitive reflection and decision-making

Many other papers, instead, focused their attention specifically on the effect of cognitive reflection on the quality of the choice and decision-making. Cognitive miser seriously impedes people's ability to make informed decision and deeply influences not only the decision-making process.

a) Thoma and colleagues ([43]) for example investigated differences in decision-making style and risk-taking among financial traders, non-trading bank employees, and people not working in finance. What they found is that traders have an higher cognitive reflection and financial risk-taking was higher in the expert groups (working in financial services) and it correlated with cognitive reflection scores (pp. 2). Financial traders have a higher propensity to inhibit the use of mental heuristics in decision-making.
b) Graffeo and colleagues found that cognitive reflection drives the quality of the consumers' decision-making process in a purchase problem solving context ([37]). Subjects with lower cognitive reflection chose the worse deal more frequently and showed a more superficial analysis and information search pattern.
c) Simonovic and colleagues ([44]) studied the importance of cognitive reflection in the Iowa Gambling Task (IGT) while manipulating the level of stress, and they found that people with higher cognitive reflection have better IGT performance, and they appear to learn more from the outcomes of their decisions even when stressed.
d) Cognitive reflection also mediates socially biased decisions. Baldi and colleagues ([45]) found that cognitive reflection is associated with a reduction of decision-making bias associated with social status so people with a higher level of cognitive reflection are less influenced by celebrity status.
e) Moritz and colleagues ([46]) showed that managers with higher cognitive reflection make more profitable decisions.

Cognitive reflection and metacognition

Other studies analyzed the specific role of cognitive reflection on metacognition because it has been demonstrated that the propensity to think analytically facilitates metacognitive monitoring during reasoning; cognitive reflection, as a measure of the propensity to engage in analytical reasoning, is related with:

a) awareness of what we are doing or what we think ([47]);
b) "faith in intuition" (FI), the individual's trust in her own intuition ([33])

even if a gender difference occurs (women reported higher FI than male subjects);

c) lower scores in cognitive reflection are related to Dunning-Kruger effect ([48]) by which the incompetent are often unable to recognize their incompetence;
d) the incapacity to discern fake news from real news ([49]);
e) CRT was used also to test a dual-processing account of another kind of metacognitive monitoring, lucid dreaming ([50]); Rizea and Malinowski demonstrated that there is no relationship between rational reflective abilities, measured with the CRT, and lucid dream frequency.

If it is true that higher cognitive reflection leads to higher metacognitive monitoring, also the opposite is true; engaging in mindfulness meditation affects analytical thought processes. Participants who listened to a mindfulness recording have higher CRT performance compared to the control group ([51]).

Cognitive reflection, God and political orientation

In the literature, belief in God is defined as a natural, automatic and intuitive behaviour that occurs in our everyday life; it has been demonstrated that people with lower cognitive reflection reported stronger belief in God ([52]) because they are more prone to anchor on intuitions (and are more prone to superstitious thinking too, [53]). Cognitive reflection also predicts political orientation: liberals have a higher cognitive reflection compared to conservatives ([54], [31]) even if other papers (e.g. [55]) demonstrate that CRT has no significant relationship with political orientation.

Cognitive reflection and cooperation

No significant association was found between CRT and the strict utilitarian and deontological response pattern, and willingness to act in the utility-optimizing manner ([56]). However, Lohse ([57]) demonstrated that people with higher cognitive reflection are more prone to cooperation in a one-shot public good game and to trust in trust game ([58]). In a bank-run game cognitive reflection predicts the use of dominant strategy only in strategic uncertainty version (where the last depositor has no information regarding the decisions of predecessors) but not in the condition with no uncertainty ([59]).

Cognitive reflection and time discounting

Concerning the field of time discounting, people with higher cognitive reflection tend to discount more weakly, as they are more resilient to imposed reference points than people with lower cognitive reflection which prefer smaller more immediate rewards (steeper discounters) in their choices, preferences, and beliefs. ([60], [61]). Also Frederick [1] found that individuals who performed well on his CRT were more likely to choose a future reward of greater value than a smaller immediate reward.

Cognitive reflection, reasoning strategies and cognitive flexibility

Ferrer and colleagues ([42]) created different versions of CRT in order to show that how the question is framed can influence the answers in terms of accuracy and response time. They presented a non-conflict version by making the intuitive impulse correct (resulting in an alignment question), shutting it down (creating a neutral question), or making it dominant (creating a heuristic question). Other studies divided the answers given by the participants into different categories ([62], [52]). Erceg and Bubic for example discussed five different ways of scoring the CRT: "a regular CRT scoring procedure (CRT-Regular), adding up the intuitive answers (CRT-Intuitive), calculating the proportion of intuitive in total incorrect answers (CRT-Proportion Intuitive), scoring only non-intuitive answers irrespective of their correctness (CRT-Reflection) and calculating the proportion of correct in total non-intuitive answers (CRT-Calculation)" ([53], pp. 381). Jelihovschi and colleagues analyzed different ways to solve the cognitive reflection test ([63]). They reported four different reasoning strategies in answering CRT: no expression (participants answered the questions without any calculation); organization (participants organized the pieces of information as they prefer), calculation (it is the high manipulation of data, participants tried to calculate the solution) and erasure (participants provided an answer that was deleted). Calculation and erasure predict higher performance. The first strategy is related to a more rational analysis of the information and the second strategy is related to cognitive flexibility. Elisa and Parris demonstrated the importance of attention in the problem solving process of CRT, analyzing the executive function deficits of the Attention Deficit and Hyperactivity Disorder (ADHD) ([64]). All the three core symptoms of ADHD (inattention, hyperactivity and impulsivity) are negatively related to performance on the CRT but inattention is the most important one for this cognitive ability.

How we analyze information, how we allocate our attention and the relation between attention and cognitive reflection, is the main topic of the third and fourth chapter where I will present the second and third paper "Getting the best deal: Effects of cognitive reflection on mental accounting of choice attributes" and "Cognitive Reflection and gaze behaviour in visual tasks" respectively. How we analyze information is important for understanding how we integrate pieces of information, which in turn, is related to the integration/segregation process underpinning the concept of "mental accounting" ([65]). What we assume is that people with lower cognitive reflection search for information and allocate their attention in a different way compared to people with higher cognitive reflection. The lower cognitive reflectors may mentally separate and not integrate information that does not belong to the same category because they are more prone to the categorization process derived by the System 1. We did not find literature that links cognitive reflection to mental accounting and our first exploration, using the eye tracker in order to measure attention is, in my view, extremely interesting. How we allocate our attention is also the main topic of the third paper in which we analyzed the link between executive functioning (in particular set-shifting and inhibitory control) and cognitive reflection in two visual tasks. We decided to study visual tasks in order to understand if the cognitive reflection acts also in the early step of perception. We started from the results shown by Del Missier and colleagues ([66]), and the suggestions of Toplak and colleagues ([8]), which analyzed the CRT performance in relation to monitoring/inhibition dimensions of executive functions. CRT was measured in order to understand if cognitive reflection can predict how participants allocate their visual attention in a visual search task ([67]) and in the antisaccade task ([68]). We chose these tasks because in both there is a prepotent behavioural response: in the visual search task, a matching bias leads participants to look at the colour mentioned in the search question ([69]), even though this occasionally leads to more searching than necessary. In the antisaccade task "participants must suppress the reflexive urge to look at a visual target that appears suddenly in the peripheral visual field and must instead look away from the target in the opposite direction [...] A crucial step involved in performing this task is the top-down inhibition of a reflexive, automatic saccade" ([68], pp. 218). For the first time in the literature, these tasks were analyzed as a function of CRT. Our goal is to test the hypothesis that people with higher cognitive reflection have a better ability to override a prepared or prepotent control over attention when such control is required (as in the antisaccade task) or simply beneficial (as in the search task) and, consequently, a better ability to switch behaviour depending on the best strategy to complete a task as fast as possible.

References

[1] Frederick S. Cognitive reflection and decision making. The Journal of Economic Perspectives. 2005;19(4):25–42.

[2] Kahneman D, Frederick S. Representativeness revisited: Attribute substitution in intuitive judgment. Heuristics and biases: The psychology of intuitive judgment. 2002;49.

[3] Evans JSB, Stanovich KE. Dual-process theories of higher cognition advancing the debate. Perspectives on psychological science. 2013;8(3):223–241.

[4] Baron J, Scott S, Fincher K, et al. Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? Journal of Applied Research in Memory and Cognition. 2015;4(3):265–284.

[5] Sinayev A, Peters E. Cognitive reflection vs. calculation in decision making. Frontiers in psychology. 2015;6:532.

[6] Travers E, Rolison JJ, Feeney A. The time course of conflict on the Cognitive Reflection Test. Cognition. 2016;150:109–118.

[7] Duncker K, Lees LS. On problem-solving. Psychological monographs. 1945;58(5):i.

[8] Toplak ME, West RF, Stanovich KE. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. Memory & Cognition. 2011;39(7):1275–1289.

[9] Sternberg RJ, Davidson JE. The nature of insight. The MIT Press; 1995.

[10] Corgnet B, Espín AM, Hernán-González R. Creativity and cognitive skills among millennials: thinking too much and creating too little. Frontiers in psychology. 2016;7:1626.

[11] Barr N, Pennycook G, Stolz JA, et al. Reasoned connections: A dual-process perspective on creative thought. Thinking & Reasoning. 2015;21(1):61–75.

[12] Corgnet B, Espín AM, Hernán-González R. Creativity and cognitive skills among millennials: thinking too much and creating too little. Frontiers in psychology. 2016;7:1626.
[13] Simon HA, Newell A. Human problem solving: The state of the theory in 1970. American Psychologist. 1971;26(2):145.
[14] Campitelli G, Gerrans P. Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. Memory & Cognition. 2014;42(3):434–447.
[15] Toplak ME, West RF, Stanovich KE. Assessing miserly information processing: An expansion of the Cognitive Reflection Test. Thinking & Reasoning. 2014;20(2):147–168.
[16] Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt impulsiveness scale. Journal of clinical psychology. 1995;51(6):768–774.
[17] Primi C, Morsanyi K, Chiesi F, et al. The development and testing of a new version of the cognitive reflection test applying item response theory. Journal of Behavioral Decision Making. 2015;.
[18] Primi C, Morsanyi K, Donati MA, et al. Item Response Theory analysis of the Cognitive Reflection Test: Testing the psychometric properties of the original scale and a newly developed 8-item version. In: Proceedings of the Annual Meeting of the Cognitive Science Society; Vol. 36; 2014.
[19] Thomson KS, Oppenheimer DM. Investigating an alternate form of the cognitive reflection test. Judgment and Decision Making. 2016;11(1):99.
[20] Liberali JM, Reyna VF, Furlan S, et al. Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. Journal of Behavioral Decision Making. 2012;25(4):361–381.
[21] Mastrogiorgio A. Commentary: Cognitive reflection vs. calculation in decision making. Frontiers in psychology. 2015;6.
[22] Goodman M, Finnegan R, Mohadjer L, et al. Literacy, Numeracy, and Problem Solving in Technology-Rich Environments among US

Adults: Results from the Program for the International Assessment of Adult Competencies 2012. First Look. NCES 2014-008. National Center for Education Statistics. 2013;.

[23] Fagerlin A, Zikmund-Fisher BJ, Ubel PA, et al. Measuring numeracy without a math test: development of the Subjective Numeracy Scale. Medical Decision Making. 2007; 27(5):672–680.

[24] Woodcock RW, McGrew KS, Mather N, et al. Woodcock-Johnson III NU tests of achievement. Rolling Meadows, IL: Riverside Publishing. 2001;.

[25] Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Medical decision making. 2001;21(1):37–44.

[26] Schwartz LM, Woloshin S, Black WC, et al. The role of numeracy in understanding the benefit of screening mammography. Annals of internal medicine. 1997;127(11):966–972.

[27] Weller JA, Dieckmann NF, Tusler M, et al. Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. Journal of Behavioral Decision Making. 2013; 26(2):198–212.

[28] Welsh M, Burns N, Delfabbro P. The cognitive reflection test: how much more than numerical ability? In: Proceedings of the Annual Meeting of the Cognitive Science Society; Vol. 35; 2013.

[29] Oechssler J, Roider A, Schmitz PW. Cognitive abilities and behavioral biases. Journal of Economic Behavior & Organization. 2009;72(1):147–152.

[30] Brañas-Garza P, Garcia-Munoz T, González RH. Cognitive effort in the beauty contest game. Journal of Economic Behavior & Organization. 2012;83(2):254–260.

[31] Albaity M, Rahman M, Shahidul I. Cognitive reflection test and behavioral biases in Malaysia. Judgment and Decision Making. 2014;9(2):148.

[32] Brañas-Garza P, Kujal P, Lenkei B. Cognitive Reflection Test: whom, how, when. Munich University Library. 2015;MPRA Paper No. 68049.

[33] Alós-Ferrer C, Hügelschäfer S. Faith in intuition and cognitive reflection. Journal of Behavioral and Experimental Economics. 2016;64:61–70.

[34] Morsanyi K, Busdraghi C, Primi C. Mathematical anxiety is linked to reduced cognitive reflection: a potential road from discomfort in the mathematics classroom to susceptibility to biases. Behavioral and Brain Functions. 2014;10(1):31.

[35] Nave G, Nadler A, Zava D, et al. Single-dose testosterone administration impairs cognitive reflection in men. Psychological science. 2017;28(10):1398–1407.

[36] Campitelli G, Gerrans P. Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. Memory & cognition. 2014;42(3):434–447.

[37] Graffeo M, Polonio L, Bonini N. Individual differences in competent consumer choice: the role of cognitive reflection and numeracy skills. Frontiers in psychology. 2015;6.

[38] Cardoso RL, de Oliveira Leite R, de Aquino ACB. The effect of cognitive reflection on the efficacy of impression management: An experimental analysis with financial analysts
. Accounting, Auditing &Accountability. 2018;Vol. 31(No. 6):1668–1690.

[39] Peters E, Fennema MG, Tiede KE. The loss-bet paradox: Actuaries, accountants, and other numerate people rate numerically inferior gambles as superior. Journal of Behavioral Decision Making. 2019;32(1):15–29.

[40] Mata F, Miranda DM, Sallum I, et al. Do general intellectual functioning and socioeconomic status account for performance on the Children's Gambling Task? Frontiers in neuroscience. 2013;7:68.

[41] Noori, M. (2016). Cognitive reflection as a predictor of susceptibility to behavioral anomalies. *Judgment and Decision Making*, *11*(1), 114.

[42] Alós-Ferrer C, Garagnani M, Hügelschäfer S. Cognitive reflection, decision biases, and response times. Frontiers in psychology. 2016;7:1402.

[43] Thoma V, White E, Panigrahi A, et al. Good thinking or gut feeling?

Cognitive reflection and intuition in traders, bankers and financial non-experts. PloS one. 2015;10(4):e0123202.

[44] Simonovic B, Stupple EJ, Gale M, et al. Stress and Risky Decision Making: Cognitive Reflection, Emotional Learning or Both. Journal of Behavioral Decision Making. 2016.

[45] Baldi PL, Iannello P, Riva S, et al. Cognitive reflection and socially biased decisions. Studia Psychologica. 2013;55(4):265.

[46] Moritz BB, Hill AV, Donohue KL. Individual differences in the newsvendor problem: Behavior and cognitive reflection. Journal of Operations Management. 2013;31(1):72–85.

[47] Jackson SA, Kleitman S, Howie P, et al. Cognitive abilities, monitoring confidence, and control thresholds explain individual differences in heuristics and biases. Frontiers in psychology. 2016;7:1559.

[48] Pennycook G, Ross RM, Koehler DJ, et al. Dunning-Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. Psychonomic bulletin & review. 2017;24(6):1774–1784.

[49] Pennycook G, Rand DG. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition. 2018;.

[50] Rizea AE, Malinowski JE. An investigation of a dual-processing hypothesis of lucid dreaming. International Journal of Dream Research. 2017;10(1):15–22.

[51] Fitzgerald CJ, Lueke AK. Mindfulness increases analytical thought and decreases just world beliefs. Current Research in Social Psychology. 2017;25:80–85.

[52] Shenhav A, Rand DG, Greene JD. Divine intuition: Cognitive style influences belief in God. Journal of Experimental Psychology: General. 2012;141(3):423.

[53] Erceg N, Bubić A. One test, five scoring procedures: different ways of approaching the cognitive reflection test. Journal of Cognitive Psychology. 2017;29(3):381–392.

[54] Deppe KD, Gonzalez FJ, Neiman JL, et al. Reflective liberals and intuitive conservatives: A look at the Cognitive Reflection Test and ideology. Judgment and Decision Making. 2015;10(4):314–332.

[55] Yilmaz O, Saribay SA. The relationship between cognitive style and political orientation depends on the measures used. 2017;.

[56] Royzman EB, Landy JF, Leeman RF. Are thoughtful people more utilitarian? CRT as a unique predictor of moral minimalism in the dilemmatic context. Cognitive science. 2015; 39(2):325–352.

[57] Lohse J. Smart or selfish-When smart guys finish nice. Journal of Behavioral and Experimental Economics. 2016;64:28–40.

[58] Corgnet B, Espín AM, Hernán-González R, et al. To trust, or not to trust: cognitive reflection in trust games. Journal of Behavioral and Experimental Economics. 2016;64:20– 27.

[59] Kiss HJ, Rodriguez-Lara I, Rosa-García A. Think twice before running! Bank runs and cognitive abilities. Journal of Behavioral and Experimental Economics. 2016;64:12–19.

[60] Shenhav A, Rand DG, Greene JD. The relationship between intertemporal choice and following the path of least resistance across choices, preferences, and beliefs. Preferences, and Beliefs (December 16, 2016). 2016.

[61] Bialek M, Sawicki P. Cognitive reflection effects on time discounting. Journal of Individual Differences. 2018;39:99–106.

[62] Pennycook G, Cheyne JA, Koehler DJ, et al. Is the cognitive reflection test a measure of both reflection and intuition? Behavior Research Methods. 2016;48(1):341–348.

[63] Jelihovschi AP, Cardoso RL, Linhares A. An analysis of the associations among cognitive impulsiveness, reasoning process, and rational decision making. Frontiers in psychology. 2018;8:2324.

[64] Elisa RN, Parris BA. The relationship between core symptoms of ADHD and the cognitive reflection test in a non-clinical sample. Cognitive neuropsychiatry. 2015;20(5):416–423.

[65] Thaler R. Toward a positive theory of consumer choice. Journal of Economic Behavior & Organization. 1980;1(1):39–60.

[66] Del Missier F, Mäntylä T, Bruin WBD. Decision-making competence, executive functioning, and general cognitive abilities. Journal of Behavioral Decision Making. 2012; 25(4):331–351.

[67] Rajsic J, Wilson DE, Pratt J. Confirmation bias in visual search. Journal of experimental psychology: human perception and performance. 2015;41(5):1353.

[68] Munoz DP, Everling S. Look away: the anti-saccade task and the voluntary control of eye movement. Nature Reviews Neuroscience. 2004;5(3):218.

[69] Rajsic J, Taylor JET, Pratt J. Out of sight, out of mind: Matching bias underlies confirmatory visual search. Attention, Perception, & Psychophysics. 2017;79(2):498–507.

2. Summary of Empirical Studies

The primary aim of the empirical studies was to examine the role of cognitive reflection on cognitive and visual biases strictly related to the consumer decision-making process. More specifically in the Chapter 2, through the paper "The role of numeracy, cognitive reflection and attentional patterns in commercial problem-solving" we investigated the role of cognitive reflection in two decision-making scenarios involving high numerical components in order to understand if only numeracy affects the accuracy of the decisions. In the Chapter 3, "Getting the best deal: The role of a topical mental accounting of multi-dimensional prices, and cognitive reflection", we analyzed the role of cognitive reflection on the integration of all decision attributes (what we called "intra-product mental accounting") and how this affects the accuracy of the choice. In the Chapter 4 "Cognitive reflection and gaze behaviour in visual tasks", we examined if cognitive reflection has a predictive power on biases in two visual tasks in order to understand if cognitive reflection plays an important role also in how we search for information in the initial phase of the decision-making process, related to perception.
The relationship between the CRT and different cognitive biases has been widely studied. The lack of control of System 2 on System 1, captured by this test, leads to the use of heuristics and biases. With these three studies we demonstrated the effect of this point in different tasks making valuable contributions to the literature (that will be explain in the following sections) and touching many aspects that have not been considered by the literature (e.g. the link between cognitive reflection and mental accounting or the relationship between cognitive reflection and some visual tasks).

Study 1

The aim of Study 1 was to understand if in two different kinds of commercial problem-solving scenarios involving high numerical components, only numeracy affects the accuracy of the decisions. Two are the main hypotheses of this study.

The first one is that numeracy affects accuracy because mathematical skill is essential in order to solve the problem but cognitive reflection also plays an important role in these kinds of tasks because it reflects an ability of the participants to perform a more reflective analysis of the information (e.g. to not be influenced by a heuristic analysis, choosing for example the option with the lower initial price). Therefore, we expect that both numeracy and cognitive reflection will predict accurate choice. The second hypothesis is related to the depth of cognitive information processing (measured by an Attentional Index related to eye movements) and its effect on the accuracy of the choice. Considering the definition and the literature about cognitive reflection, we expect that people with lower cognitive reflection analyze information more superficially, with use of heuristics and biases, and this leads to a worse choice performance.

Study 2

How we allocate our attention and how we integrate information is an important aspect that impacts the decision-making process and consequently the choice. In this paper we study the role of cognitive reflection on the mental accounting of choice attributes of an offered product/service package, and its impact on the accuracy of choice. Firstly, we presented the new concept of “intra-product mental accounting” and secondly we demonstrated that people with lower cognitive reflection are more prone to segregate information that does not belong to the same category of the other pieces of information. Participants were presented with two websites selling the same package, and asked to choose one. One of the two packages was always associated to a lower total price than the other. The two alternative packages were described by the same choice attributes. Some attributes were easily grouped together (similar attributes) whereas another attribute was not (dissimilar attribute). Gaze behaviour and final choice were recorded.

Study 3

In this study we investigated if cognitive reflection also affects an initial phase of the decision-making process, the visual search. We studied how participants allocate their attention in two different visual tasks (confirmation bias in visual task and antisaccade task) that involve overriding a prepotent response and we analyzed eye movements in order to understand if people with lower cognitive reflection use more biases and are not able to override prepotent responses. Important aspects of attention and information search were discussed and we have drawn important conclusions.

Appendix A - Eye Link 1000 Plus

In all of the papers reported in this thesis, we used a really powerful tool: EyeLink 1000 Plus Binocular Tower Mount, the most versatile solution for eye and gaze monitoring available. Eye tracking is growing in popularity amongst researchers from a whole host of different fields that try to analyze the attention of the people using empirical investigation. Specialized skills are required in order to use this tool and to analyze the data collected using this instrument. In this section I will report the main concepts related to eye tracking research.

2.1. Fundamental principles and technical information

What we expect from a tool is for it to provide an accurate and precise measure of the construct under examination. We want it to capture the same specific point that the participant is looking at. Precision, or the other side of the coin of variability, is a description of the random errors present in the measures of our measurement system, the statistical variability. The accuracy, and so the measure of statistical bias, is a description of the systematic errors. A measurement system can be accurate but not precise, precise but not accurate, neither, or both. In an eye tracking experiment, both these characteristics of the measurement system are crucial to correctly and exactly record what the participant is perceiving. There are many eye trackers on the market (e.g. Tobii TX300, SMI RED 250/500, Tobii T60, Eye Link 1000, LC Technologies Eye Follower), mobile and static, and each of them have strengths and weaknesses that act on the precision and accuracy of the measurement system. We have pros and cons for all the three types of the eye tracking: glasses, a mobile type (a), and a fixed type, divided into the remote eye tracking (b), with a imaginary box where the head of the participant can move, and the tower mount eye tracker (c) where the head of the participant is fixed by a chin-rest, a forehead-rest, or both. Depending on the type of research, you will need specific features of the measurement system and so you will deal with different issues. The Eye Link 1000 Plus Binocular Tower Mount provides highly accurate and precise monocular or binocular data acquisition at up to 2000 Hz (the highest sampling frequencies available) using a chin and forehead rest; it is a very powerful tool that, if used correctly, can provide a high level of research. It collects data with low variable error, so with low root mean square deviation and consequently high precision, a small systematic error, high accuracy and almost no data loss. It incorporates the camera and illuminator housing within a combined chin and forehead rest via infrared reflective mirror. The Tower Mount affords the largest field of view of all mounting systems. The average accuracy measured with real eye fixations at multiple screen positions on a per subject basis, is down to 0.15◦, with a blink and occlusion recovery of 0.5 ms and a spatial resolution measured with an artificial eye, of 0.01◦. The gaze tracking range is 60◦ horizontally and 40◦ vertically and the allowed head movements for +/- 25 mm horizontal or vertical.

However, even if this is a really professional tool for the gaza detection, there are several procedures necessary to achieve the precision and accuracy that this tool can reach.

(1) First and foremost, to define a specific research question, as we said above, is necessary for three main reasons: to prevent an inappropriate use of eye tracking in experiments that do not need eye tracking in order to answer to the research question; to better understand the best eye tracker type for our research and to lead an appropriate use of the Area of Interest (AOI). This last point will be analyzed in the next section.

(2) Secondly, we need a calibration procedure as accurate and precise as possible. This is an important and a fundamental process that we have in order to align what the tool measures with what the participant sees. We could have the most powerful tool, but if we calibrate in the wrong way, we will most likely collect inaccurate responses. The underlying concept which has forms the basis for eye tracking measures is that the human vision is divided into two main parts: a small central area with very high resolution, called foveal vision, and the vast majority of the visual field with poor resolution, called peripheral vision. What we want to do with the initial calibration and validation measures is to align the position of the

stimulus with the foveal vision, in order to collect precise and accurate data about where participants are looking. Bad calibration procedures may cause systematic and random errors (for an overview on this topic see [1]). Other reasons could be, for example, the mascara on the eyes of the subject, the glasses, contact lenses, dirty mirrors or bad lighting in the lab.

A small error rate is physiological because we have a really specialized human-machine interaction and because the human eye is subject to small, rapid, constant and involuntary eye movements, called microsaccade with an amplitude that varies from 2 to 120 arcminutes. For these reasons, most of the software that collects and analyzes eye tracking data files have embedded algorithms for dealing with noise detection, fixation, and saccade recognition issues related to measurement errors. There are also many event detection algorithms (factories).

The function of eye movements and consequently what an eye tracker measures and studies, is to direct the eye to a new part of the visual field and to keep parts of the world fixed on the retina, with the foveal vision. The first function is associated with the saccade, a fast and most of the time voluntary movement from one fixation to the next. Sometimes it is also a reflex because it is related with the automatic nervous system and more specifically with the fight or flight response that leads a person to look at a potential harmful event or attack that occurs close to her [2], [3]. Instead, the smooth pursuit, the vestibulo ocular reflex (VOR), the optokinetic reflex (OKR) and the fixation have the common goal of keeping the world fixed on the retina.

(a) The smooth pursuit eye movements are slow and both voluntary or reflex and they allow the eyes to follow a moving object;
(b) the eye also has a tendency to move after a saccade and before it fixates on an object, this movement is known as glissade;
(c) the VOR is a fast reflex that stabilize the images on the retina during head movements producing eye movements in the opposite direction of the head movements;
(d) the OKR is an eye movement that allows to follow a moving object that moves out of the field of vision, to return then to the position when it first saw the object; a fixation that occurs when we maintain the visual

gaze on a single location.

We did not consider the millisecond duration of each element mentioned above because, as we will see in the section "Measures" we should define the properties of these eye movements or fixation, considering our research field and the features of the stimulus. As presented in the paper "Is the eye-movement field confused about fixations and saccades?", we also have to consider that different definitions may be used across different fields and within one field. Confusion could be due to three possible sources: (i) different frames of reference, (ii) functional versus oculomotor versus computational definitions of fixations and saccades, and (iii) event classification in the eye-tracker signal [4]. I will get into the nature of measurements in the specific section. However, it is important to understand the movements and fixations related to the tool that we used.

- With the tower eye tracker (head fixed and with the eye tracker output signal, the point of regard to pixels looked at, in the world fixed coordinates), with a static stimulus we can measure only fixations and saccades; with a dynamic stimulus we can measure also smooth pursuits and OKRs;
- with the remote eye tracker (head free and with the eye tracker output signal, the point of regard to pixels looked at, in the world fixed coordinates), with a static stimulus we can measure only fixations and saccades; with a dynamic stimulus we can measure also smooth pursuits and OKRs;
- with the glasses (head free and with the eye tracker output signal, the point of regard to pixels looked at, in the scene camera image, in head-fixed coordinates) we have only dynamic stimulus and we are able to measure all of the movements, saccades, smooth pursuits, OKRs and VORs.

A difficult issue to solve for software that analyzes eye tracking data, is to disentangle saccades from smooth pursuits. It is not only a matter of velocity of the movement because also a smooth pursuit could be as fast as a saccade. For example, the velocity-based adaptive algorithm has been proposed in order to allow efficient saccades detection but also a more recent version, the

Sacc/SP algorithm [5].
In all of the studies presented in this thesis we used a tower eye tracker (Eye Link, 1000 Plus) with static stimulus. It is always better to use the simplest set up to answer the research question.

2.2. Area of Interest

The Areas of Interest, AOIs, are regions in the stimulus that are interesting for the research question, and they are used to quantify whether and how many participants looked at the particular regions. They are also known under different names, such as IAs (interest area), ROIs (regions of interests) and Zones.

First of all it is necessary to keep the research question in mind when we draw the AOIs because different research questions require a different subdivision of the interest area. Take, for example, the first fixation of a photo representing a woman and a man. The AOIs to test the hypothesis that the first fixation will be on the eye of the two people, is completely different to the AOIs to test the hypothesis that the first fixation will be on the woman rather than the man. As we can imagine, the interest areas in the first case will divide the eyes of the two people to the other part of the photo. In the second case, instead, they will divide the woman's figure to the man's figure. Setting up different AOIs will lead to different analyses to test different hypotheses. If you do not think ahead to the AOIs it is possible that you will not be able to disentangle if the participant focused more on a specific part or another. Different hypotheses need different AOIs and different AOIs allow testing of different hypotheses. This does not mean that after a first analysis based on a specific subdivision of the space you cannot reset the AOIs and manipulate them in order to test another hypothesis. If you draw your AOIs after data recording, you are forming post-hoc hypothesis, based on the results while inspecting the data. If you have a hypothesis about statistics of certain gaze locations then the AOIs you draw are implicitly part of the hypothesis, so if you alter your AOIs you alter your hypothesis and your statistics. They are part of your hypothesis. It is important to keep the research question in mind when drawing AOIs because they are the main tools for further analysis of eye-movement data. They allow to define and detect AOI hits, dwells, transitions and the AOI total skip.

There are several advices to consider before drawing the AOIs but the main advice regards the dimension of the AOIs and their location. It is important not to make AOIs too small; the minimal AOI size is limited by the precision and accuracy of the eye tracker. An error of 0.5° means a circle error with a radius of 0.5 cm on a screen placed at 57 cm. If you AOIs are too small, a possible error leads to consider a fixation in the AOI out of the interest area or the other way around. It is important to draw the AOIs as large as possible considering the research hypothesis because it allows to capture also the fixation that are subject to error.

There are different production methods for drawing interest areas. They are created using a tool for spatial segmentation usually supplied in the analysis software for the eye tracker, called AOI editor or they can create customize interest areas directly from the programming software used to design the experiment (e.g. MATLAB) indicating the coordinates of the interest area. In the first case you can have hand drawn AOIs, pre-set shapes drawn AOIs, or AOIs automatically generated by the stimulus. You can base the drawing of the AOIs on different criteria (semantic AOI e.g. [6], distributed AOI used most the time in the marketing research e.g. [7] and gridded meaningless AOI for a string representation of the scan path and dwell map [8]). AOIs can be based on static images or dynamic depending on the type of the stimuli.

The basic AOI events, as we said above, are AOI hits, dwells and transitions.

- The AOI hit is "the most primitive AOI event, which states for a raw sample or a fixation that its coordinate value is inside the AOI, it underlies all raw AOI measures, including those based on fairly complex representations like proportion over time graphs" [9].

- The dwell, or gaze or glance, is a visit to an AOI from entry to exit. It collapses all the information about a specific AOI, giving for example the duration, the starting point and the ending point of a gaze into that interest area. Dwells can be represented with dwell map or hit map. The first one is used most of the time with the gridded AOIs and it is a table with the dwell time of the all the single cells in the grid. The heat map uses different colours to represent the dwell duration, instead of representing this duration in numbers (Figure 6): in a scale from red to green where red represents the longer duration of the dwell and the green the shortest one, with varying levels in between. That is why the heat maps are colour-coded. Both these kinds of maps can be used to quickly understand the visual hierarchy of the elements of the stimulus in order to make, for example, "informed decisions about how to optimize the placement of screen elements" [10].

- The transition, or gaze shift, is the movement from one AOI to another. You can also have a return, or revisit, that is a transition to an AOI already visited; or a within AOI transition, when there is a saccade within an AOI.

Figure 6. A heat map taken from [11]. From this picture we can understand that people did not look at the tiny images on the Panasonic site.

This would be considered a proper transition because the term transition implies an eye movement from one AOI to another. We have another conceptual problem when the transition is not linear from an interest area to another but spaced by a fixation out of the first AOI before entering to the other. We could have also another situation when an AOI is skipped in the transition from an AOI to another (AOI first skip). When an AOI is not looked for the total duration of the trial, it is called AOI total skip (e.g. an advertisement on a newspaper).

When you consider a transition between two AOIs you are automatically considering this process over time. Adding the variable "time" allows you to analyze the movements among AOIs or fixation of a specific area. The AOI string is a sequence of either fixation-based AOI hits or dwells that consider occurrence during time. As mentioned in [9] there are at least three different varieties:

- A string is given by all the fixations on the different AOIs expressed in letters (e.g. LLHGGMCL means that the first two fixations were in the L interest area, the third in the H interest area and so on and so forth). If there are two fixations they both will be expressed.
- A compressed string that considers only dwells, so without repetition (the previous example will become LHGMCL).
- A string that considers only the first dwell in every single AOI (the previous example will become LHGMC); the maximal string length may be maximum the number of the AOIs presented in the stimulus.

The transition matrix and the use of the Markov model allow to use probabilistic model describing or modelling the data. However, there are different ways to analyze the data; we will see them in the next section.

2.3. Measures

The main attribute measures that we can analyze from an eye tracking data file are the location of the eye gaze, the duration of a fixation to a particular area and the movement from one fixation to another. The location of a fixation is represented by the x and y coordinates on a grid where the stimulus can be placed in order to understand the elements presented on the screen. Usually a fixation is extremely short, between 70 and 600 milliseconds and it is recorded on the eye tracking data file depicted by a detection algorithm of the eye tracking software. The information related to the fixation is extremely important but we should pay attention to what we can infer from that. A fixation does not necessarily imply that the brain is processing the information and it is true also the opposite: information that is not in the foveal area is not necessarily not analyzed (it could be analyzed). Anyway, in the literature is assumed that fixations depict the focus of attention on an object even if using an eye tracker we cannot confidently infer anything about a person's analysis because the fixation may not accurately and precisely represents cognitive processing; we have to be very careful about inferring the type of processing starting from a fixation. The duration of the fixation is another important cue that can help us to understand if the gaze is related to the attention. A clustering of a number of fixations in a particular region and their duration can tell us that it is very likely that the user is paying particular attention to that specific area. The problem is to understand why she is looking at that visual element. This could mean that there is a high cognitive load, she is processing the information more in depth, or she is confused by this element.

During fixations the eye is still, even if some micro-movements like drifts, tremors and micro-saccades can occur; saccades are eye movements between fixations ([12]). In the literature, two are the main eye tracking data analysis: fixation analysis (1) and saccade analysis (2). A fixation analysis depicts the focus of attention and it is the most often used measure in eye tracking research ([13]). It is divided into two kinds of analysis: (1.1) fixation duration (given by the focus of attention parameters: "dwell time" and "total dwell time") and (1.2) fixation counts.

(1.1) Dwell time, also known as glance duration, visit duration, glance or gaze ([12], [9]), is defined as a sum of all fixation durations during a dwell in an area of interest (AOI) on the stimulus. It has its own duration, starting point and ending point. Longer fixations are associated to deep information processing ([14], [15]). However, why the person is processing that information in that way it is difficult to infer (e.g. it could be due to the difficulty of the task, or for the significance of that specific area of interest).

(1.2) Fixation counts often represent general understanding of the focus of attention. There is a negative relationship between fixations duration and fixation counts: the greater the number of fixations, the lower their duration and vice versa ([12], p. 412). When an object is meant to draw attention in the stimulus, this increases fixation counts in that AOI. Fixation counts provide some information about processing of information (e.g. fewer fixations could mean faster processing of information from the stimulus; level of expertise, experts have fewer fixations compared to novices and only on relevant areas).

Eye Link 1000 plus also records data in addition to fixations, such as saccades, blinks, glissades, micro-saccades and pupil size. The visual hierarchy of a scene is given by a saccades analysis.

(2) The number, proportion, rate and direction of the saccade provide information about how people analyze the information. There are several questions that this analysis tried to answer (e.g. "In what direction did the eye move?"; "How far did the eye move?"; "For how long did the eye move?") using the direction of the saccades in degrees.

Using saccadic direction, we know from the literature that:

(a) most saccades are aligned with the picture horizon ([16]) and horizontal saccades are more frequent compared to vertical and oblique saccades (oblique saccades are the rarest) ([17]; [9], [18]).

(b) Saccadic velocity is affected by the age, the level of sleepiness, the difficulty of the task, mental and neurological disorder, drugs and alcohol.

(c) Saccadic velocity decreases in people older than 75 years (compared to participant younger than 43), in tired people with low level of vigilance, in people with melancholia or Alzheimer's disease.

(d) Conversely, saccadic velocity increases during the REM sleep phase and in people under the effect of drug or alcohol.

(e) Saccadic velocity increases as the difficulty of the task increases and decreases with increasing time for a task ([9]).

An analysis based on saccades allows an accounting of the transition number, proportion and rate. Transitions are movements between areas of interest; as described by Holmqvist and colleagues ([9], pp.422) the number of transitions between two areas has been used as a measure to evaluate "the importance of an area of interest, when an area is so important that it must be more or less continuously monitored, transitions rates drop". The number of returns to an area, also called "re-fixation", is informative too because it represents how much the area is informative, and it is linked to the need to refresh working memory.

Eye tracking has been applied to numerous fields, including cognitive psychology, psycholinguistics, gaze signaling in social interaction, perception of art and film, visual search, game theory, marketing, ergonomics, education and developmental psychology, industrial engineering, human factor, human-computer interaction, computer science and user experience (ux), in order to study visual attention on a stimulus. Eye trackers are also used to study the behaviour of a participant in a specific situation, in real life, such as a supermarket, or in a virtual world (e.g. driver fatigue [19], or the effect of a mobile use during driving [20]). Following the subdivision of Duchowski ([21]) about the eye tracking applications we could have four main fields of research in eye tracking studies.

(1) The first one is related to neuroscience and psychology where the neurophysiological investigation is associated to perception issues, such as illusory contours, scene perception and visual search.
(2) The second is more related to industrial engineering and human factors where researchers study the human behaviour linked to the eye movements in potentially dangerous situations (e.g. aviation, driving and visual inspection, such as drug inspection or medical X-ray inspection).
(3) The third is related to computer science and human computer interaction; researchers focus on the interfaces between people (users) and computers.
(4) The last main field is related to marketing and advertising (e.g., web pages design and how it drives the consumer behaviour, print advertising, television enhancements, product layout design).

Eye tracking is now being widely used in a combination of the second and the third point to understand how users interact with a variety of devices, software, websites, video games, etc. From this field, most of the time, scientists want to analyze, for example, the time spent on a web page or on an advertising, what drives the attention in the first 10 seconds, which are the areas that are most viewed, if there are some areas that are never looked, what are the attributes of the first fixation. It is an extensively analyzed field with hundreds publications.

In our studies we did not use a self-report of where the participants looked because in the literature it was determined to be an unreliable measure ([22], [23]). We used dwell time and dwell counts to study how people with different level of cognitive reflection analyze pieces of information. In the next section we will summarize the three studies carried out with the use of eye tracker.

References

[1] Nyström M, Andersson R, Holmqvist K, et al. The influence of calibration method and eye physiology on eyetracking data quality. Behavior research methods. 2013;45(1):272– 288.

[2] Jansen AS, Nguyen XV, Karpitskiy V, et al. Central command neurons of the sympathetic nervous system: basis of the fight-or-flight response. Science. 1995;270(5236):644–646.

[3] Gross JJ. Sharpening the focus: Emotion regulation, arousal, and social competence. Psychological Inquiry. 1998;9(4):287–290.

[4] Hessels RS, Niehorster DC, Nyström M, et al. Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers. Royal Society open science. 2018; 5(8):180502.

[5] Larsson L, Nyström M, Stridh M. Detection of saccades and postsaccadic oscillations in the presence of smooth pursuit. IEEE Transactions on Biomedical Engineering. 2013; 60(9):2484–2493.

[6] Võ MLH, Smith TJ, Mital PK, et al. Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. Journal of vision. 2012;12(13):3–3.

[7] Holmberg N, Sandberg H, Holmqvist K. Advert saliency distracts children's visual attention during task-oriented internet use. Frontiers in psychology. 2014;5:51.

[8] Steins G, Wittrock K, Haep A. Contents of classroom management: what is necessary, what is possible, how is it done at school? Creative Education. 2015;6(19):2045.

[9] Holmqvist K, Nyström M, Andersson R, et al. Eye tracking: A comprehensive guide to methods and measures. OUP Oxford; 2011.

[10] Bergstrom JR, Schall A. Eye tracking in user experience design. Elsevier; 2014.

[11] Nielsen J, Pernice K. Eyetracking web usability. New Riders; 2010.

[12] Andersson J, Gustafsson S, Engwall E, et al. Portable eye tracking device .

2015.

[13] Cortina KS, Miller KF, McKenzie R, et al. Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with expert and novice teachers. International Journal of Science and Mathematics Education. 2015;13(2):389–403.

[14] Velichkovsky BM. From levels of processing to stratification of cognition Converging evidence from three domains of research. Stratification in cognition and consciousness. 1999; 15:203.

[15] Velichkovsky BM. Heterarchy of cognition: The depths and the highs of a framework for memory research. Memory. 2002;10(5-6):405–419.

[16] Foulsham T, Kingstone A, Underwood G. Turning the world around: Patterns in saccade direction vary with picture orientation. Vision research. 2008;48(17):1777–1790.

[17] Tatler BW, Vincent BT, et al. Systematic tendencies in scene viewing. Journal of Eye Movement Research. 2008;2(2):1–18.

[18] Lee SP, Badler JB, Badler NI. Eyes alive. In: ACM transactions on graphics (TOG); Vol. 21; 2002. p. 637–644.

[19] Ingre M, T, Peters B, et al. Subjective sleepiness, simulated driving performance and blink duration: examining individual differences. Journal of sleep research. 2006;15(1):47– 53.

[20] Balk SA, Moore KS, Steele JE, et al. Mobile phone use in a driving simulation task: Differences in eye movements. J Vis. 2006;6(6):872.

[21] Duchowski AT. Eye tracking methodology. Theory and practice. 2007;328.

[22] Guan Z, Lee S, Cuddihy E, et al. The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In: Proceedings of the SIGCHI conference on Human Factors in computing systems; 2006. p. 1253–1262.

[23] Cutrell E, Guan Z. What are you looking for?: an eye-tracking study of information usage in web search. In: Proceedings of the SIGCHI conference on Human factors in computing systems; 2007. p. 407–416.

STUDY 1

CHAPTER 2

The role of numeracy, cognitive reflection and attentional patterns in commercial problem-solving

ABSTRACT

Many everyday commercial decisions deal with different situations and amounts of information. Most of the time the consumer manages and compares numerical information. The aim of this paper is to investigate the cognitive skills that drive this choice process and the role of attentional patterns, cognitive impulsivity and numeracy on the accuracy of the choice. The experiment was recorded by the eye tracker, Eye Link 1000. Results indicated that patterns of attention during decision-making predict the accuracy and are associated with cognitive reflection.

Keywords: *numeracy, cognitive reflection, eye-tracking, economic decision-making*.

1. Introduction

Individual differences in cognitive abilities and skills play an important role in both problem-solving and decision-making [1]. With this paper we want to shed light on the role of cognitive reflection during the choice process in order to highlight the importance of considering different levels of consumer vulnerabilities. As will be developed in the Discussion section this analysis is important not only as a research question but also for its practical implications. To target consumer policy measures in order to safeguard vulnerable consumers, it is essential to know in which aspects they are vulnerable. More deeply we want to understand if, depending on cognitive reflection and numeracy, people show different attentional pattern. How we analyze information could affect the accuracy of the choice and it could be driven by these individual differences.

1.1. Numeracy

Numeracy is a cognitive skill recently and widely studied for its characteristics and for its predictive power in different fields. It has been defined as a quantitative literacy or mathematical literacy and there are several definitions in literature related to aspects of this construct. For example, in Kirsch [2] numeracy has been defined as "the knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials, such as balancing a checkbook, figuring out a tip, completing an order form, or determining the amount of interest on a loan from an advertisement". In the Programme for the International Assessment of Adult Competencies (PIAAC), designed and coordinated by the Organization for Economic Cooperation and Development (OECD), this issue was addressed in its own right as the ability to access, use, interpret and communicate mathematical information in order to deal with different situations in adult life. Math is an important component of everyday situations (e.g. shopping, cooking, interpreting information in the media, playing an instrument, financial transaction, etc.) and, therefore, numeracy can play an important role in all of these activities. Low levels of numeracy limit access to educational training and jobs and it can hinder performance and productivity [3]. Fagerlin et al. (2007) developed the subjective numeracy 6-point Likert scale [4], a subjective measure regarding people's beliefs about their skill in math, and people's preferences regarding the presentation of numerical information. One of the most used numeracy scales was developed by Lipkus and colleagues (2001). It is an 11-item test [5], which extends previous work by Schwartz et al. (1997) [6]. A quick test for statistical numeracy and risk literacy was also developed, The Berlin Numeracy Test [7] and the abbreviated numeracy scale by Weller et al. in 2013 [8] was presented as a shorter, psychometrically-improved measure of numeracy based on Lipkus' scale and Peters' scale [9]. As a test of mathematical and probabilistic reasoning we used the 8-item numeracy scale developed by

Weller et al. (2013) [8] (see Appendix A). This scale consisted of eight items, five from the original Lipkus et al. scale two from the CRT scale by Frederick, and one of the Peters et al. items [9].

1.2. Cognitive Reflection

Cognitive Reflection Test (CRT), initially studied by Frederick (2005) [10], is a predictor of rational thinking and reasoning ability [11], [12]; decision-making skills, time and risk preference [10] and thinking dispositions [13] which in turn are related to economic behaviours and decisions. Toplak et al. (2011) [11] explored the predictive properties of the CRT in a much wider range of the heuristics and biases tasks, and judgment and decision-making tasks. In Noori's study (2016), [14] people with lower cognitive reflection were significantly more likely to exhibit the illusion of control, overconfidence, conjunction and base rate fallacy, and conservatism. Graffeo et al. (2015) [15] found that cognitive reflection drives the quality of the consumers' decision-making process in a purchase problem-solving context. Subjects with lower cognitive reflection choose the worse deal more frequently and show a more superficial analysis and information search patterns. Simonovic et al. (2016) [16] studied the importance of cognitive reflection in the Iowa Gambling Task (IGT) where the level of stress was also manipulated, and they found that people with higher cognitive reflection have better IGT performance and that they appear to learn more from the outcomes of their decisions even when stressed.

In the present study, we consider and compare three measurements of cognitive reflection in addition to Frederick (2005) (see Appendix B). The original CRT was a three-item test, now widely known, and many other tests have been developed. Many cognitive reflection tests have been developed for two main reasons, to cope with the spread of the correct answers of Frederick's (2005) and to address numeracy confounding in the cognitive reflection test. We used three tests [12], [17], [18]. Toplak et al. (2014) added four items (CRT4) to the original three-item test (CRT3), to have a seven-item test (CRT7). Primi et al. (2015) investigated not only the psychometric properties of the original CRT and its reliability but also a new version of it adding three items in a test that they called CRT long (CRT-L). Thomson and Oppenheimer (2016) developed a four-item test (CRT-2) in order to increase the number of questions in the test and to address numeracy confounding. Many studies have found correlations between numeracy and CRT [19], [20] , [21] suggesting that the CRT is a measure of both cognitive reflection and numeracy. For the first time all these tests were used in the same experiment. The full set of questions can be found in Appendix A and Appendix B.

2. Experiment

The goal of this study is to investigate the role of numeracy and cognitive reflection on the quality of consumer choice in two decision scenarios. More in depth we want to analyze the importance of the Attentional Index (which we will explain later) on the accuracy of the choice and how the two individual differences are linked with this index.

The two decision scenarios present a hypothetical comparison between two shops that sell the same product with two different initial prices and discounts; and a hypothetical comparison between two currency exchange services that change money with two monetary exchanges and fees. In each decision scenarios, one of the two options is more convenient than the other. We designed these scenarios with a strong numerical component in order to force people to make a considerable cognitive effort to solve the choice task. Given this context, in this experiment CRT is used as a proxy for the willingness and/or ability to engage in a more complex, effortful, and accurate strategy while we are less interested in the CRT as a proxy for inhibition/stopping ability.

The mental calculation needed to choose the dominant option requires two attentional aspects: 1) the participant maintains the attention either in the price (monetary exchange) area or in the discount (fee) area (so that only one of the values must be kept in working memory during the calculation phase); long fixations are expected because the mental calculation of the final price is a complex cognitive operation that is composed of distinct functional processes that require deep information processing. Longer fixations are commonly associated with deep information processing [22], [23], [24]. Computational load and fixation length are related and "fixation duration can be used to provide insights into cognitive processes [. . .] cognitive processes comprising conscious mathematical steps of information integration should go along with long fixations, whereas scanning and automatic processes should produce mainly short fixations" [25].

We calculated an index that expresses how much the participant's analysis is in line with what is to be expected by an individual who tries to calculate the final price (or the final amount of money) for each of the two shops (exchange offices).

Hypothesis 1

We suppose that the participants' analysis of the information, expressed by the Attentional Index, predicts the performance in the task.

H1: We expect that participants' analysis of the information, expressed by the Attentional Index, predicts the accuracy of the choice. As we said above, calculations need a deeper analysis of the information, so we expect that the Attentional Index predicts the accuracy of the choice.

Hypothesis 2

H2.1: We expect that cognitive reflection predicts the depth of cognitive processing and the type of analysis carried out by the participants during the problem-solving phase, measured by this Attentional Index. In regard to the definition of cognitive reflection, we expect that people with lower cognitive reflection analyze information more superficially.

H2.2: We also expect that numeracy is not strictly linked to the Attentional Index and a potential relationship between these two variables could be totally explained by the cognitive reflection mediation.

Hypothesis 3

We expect a partial mediation effect of the Attentional Index on the CRT predicting accuracy. In order to test this hypothesis we have to control if:

1. the Attentional Index predicts the accuracy (information obtained from the hypothesis 1);
2. The CRT is a predictor of the choice accuracy;
3. The CRT predicts the Attentional Index.;
3. conducting a multiple regression analysis with CRT and Attentional Index predicting accuracy the effect of the Attentional Index is significant and the effect of CRT is no longer significant (full mediation) or it has decreased (partial mediation).

Considering the literature reviewed earlier, we think that cognitive plays an important role in these kinds of tasks because it reflects an ability of the participants to perform a more reflective analysis of the information (e.g. to not be influenced by a heuristic analysis, choosing for example the option with the lower initial price). Therefore, we expect that cognitive reflection will predict accurate choice through its power on attentional patterns.

3. Method

Fifty-three participants were recruited from the University of Trento. We collected as many participants as our resources allowed to increase statistical power. We excluded 1 participant for a missing-data problem. They received an amount of money proportional to their performance. The experiment was divided into two main parts. In the first part, we asked participants to make decisions in two different types of scenarios. We used problems that have an optimal solution that allows us to investigate the role of cognitive skills on the quality of choice. We had two different scenarios that we will call product and exchange.

3.1. Scenarios

In the product scenarios, participants saw forty-four different items. Each item describes the same product for sale in two shops with different initial prices and discounts (see Figure 1). Subjects were asked to choose the shop where they would prefer to buy the product. Each item presents a different product (e.g., iPhone, DVD player, sofa...). In the exchange scenario, we presented 20 items in a context that describes an identical currency in two currency exchange shops with different money values and fees. Participants were asked to choose the shop where they would prefer to change their money. Each scenario uses a different currency; from the two offices, they received two pieces of information, monetary exchange, and fees (see Figure 2).

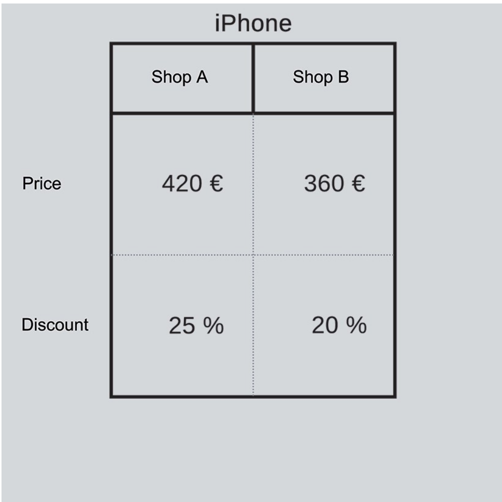

Figure 1. Product scenario.

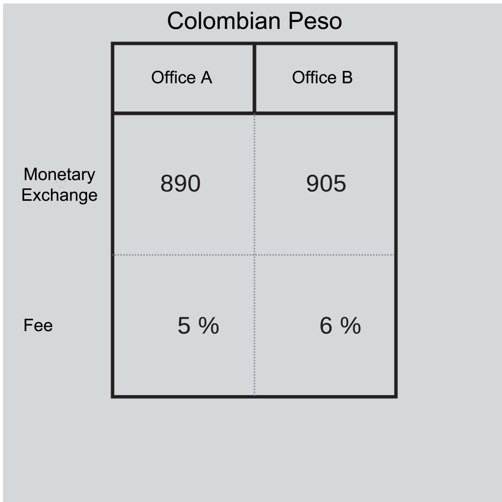

Figure 2. Exchange scenario.

After participants took a practice test, we explained that the rule for the money reward depended on their performance in the tasks.

The experiment is a 2x2 factorial design with the two scenarios as the within subject factor and the position of the price (or monetary exchange) as the between factor (it can be on the first row or on the second). In the Product scenario, as in the Exchange Scenario, the option with the lowest initial price (or the higher monetary exchange) is the dominant one 50% of the time. During the decision-making part, we recorded the position of the participants' gaze using an Eye link 1000, in order to investigate the search and analysis of participants' information processing.

4. Results

As a first analysis we created a correlogram (Figure 3) that represents the correlation matrix (Table 1) with the scores of the tests for the individual differences (numeracy and CRT) measured in the experiment in order to understand the relationship among the variables that we will use in the next analysis.

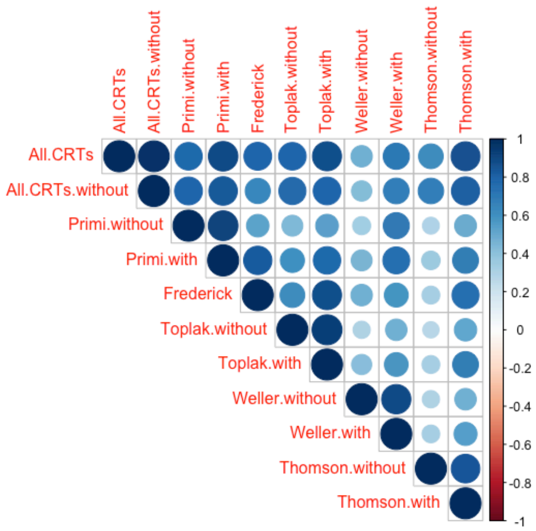

Figure 3. Correlogram, a graphical display of the correlation matrix, highlighting the most correlated variables in a data table. Positive correlations are displayed in blue and negative correlations in red colour. Colour intensity and the size of the circle are proportional to the correlation coefficients. In the right side of the correlogram, the legend colour shows the correlation coefficients and the corresponding colours.

	Frederick	Primi.without	Primi.with	Thomson.without	Thomson.with	Toplak.without	Toplak.with	Weller.without	Weller.with
Frederick	1.00	0.52	0.83	0.33	0.76	0.63	0.87	0.48	0.59
Primi.without	0.52	1.00	0.91	0.29	0.49	0.45	0.53	0.33	0.70
Primi.with	0.83	0.91	1.00	0.35	0.69	0.60	0.77	0.45	0.74
Thomson.without	0.33	0.29	0.35	1.00	0.84	0.27	0.33	0.31	0.32
Thomson.with	0.76	0.49	0.69	0.84	1.00	0.51	0.69	0.48	0.55
Toplak.without	0.63	0.45	0.60	0.27	0.51	1.00	0.93	0.30	0.47
Toplak.with	0.87	0.53	0.77	0.33	0.69	0.93	1.00	0.42	0.58
Weller.without	0.48	0.33	0.45	0.31	0.48	0.30	0.42	1.00	0.88
Weller.with	0.59	0.70	0.74	0.32	0.55	0.47	0.58	0.88	1.00
All.CRTs	0.80	0.78	0.90	0.63	0.86	0.78	0.87	0.47	0.70
All.CRTs.without	0.65	0.80	0.84	0.67	0.82	0.76	0.79	0.42	0.68

Table 1. Correlation matrix of all the main variables measured in the experiment.

As we mentioned above, the 8-item numeracy scale developed by Weller et al. [8] consisted of eight items, five from the original Lipkus et al. scale two from the CRT scale by Frederick, and one of the Peters et al. items [9]. We mentioned “Weller with” when we consider the integral version of the numeracy scale by Weller and “Weller without” when we remove the two Frederick’s items in order to avoid a higher correlation with the CRT scales. Indeed, as we can see from the table, the “Weller without” version is less correlated to the CRT scales. We considered also all the CRT scales without the three Frederick’s questions in order to analyze them with “Weller with” in the same regression. Anyway, since the CRT scales are designed in order to measure cognitive reflection we conceptually prefer the integral version of the CRT scales (the combination of the following tests: Frederick, Primi et al., Toplak et. al., Thomson and Oppenheimer) combined with “Weller without” version (the Weller numeracy test without the two Frederick’s questions).

4.1. Choices

We ran both generalized linear mixed models and linear mixed models (depending on the nature of the dependent variable), which incorporated fixed-effects parameters and random effects, to evaluate the hypothesis above mentioned. As widely discussed in the article "Fitting Linear Mixed-Effects Models Using lme4" [26], we fit random effects associated with non-nested grouping factors. Such models are common in item response theory, where subject and item factors are fully crossed. We ran models where the intercept varying among subjects and trials.

4.2. Hypothesis 1

We calculated an index that expresses how much the participant’s analysis is in line with what is to be expected by an individual who tries to calculate the final price (or the final amount of money) for each of the two shops (exchange offices). We considered (AIi) as a dependent variable and the

CRT composite measure as an independent variable. The index is defined as follows.

- The subject i has a specific Attentional Index, AIi. It indicates the total time that i remains focused on the price (or monetary exchange) areas, which we denote with Ti,p, compared to the discount (fee) areas, Ti,d. This index is given by the following formula:

$$AIi = (Ti,p - Ti,d)/(Ti,p + Ti,d)$$

The total time spent fixating on the price minus the total time spent fixating on the discount divided by the total time fixating on the price plus the total time fixating on the discount.

We calculated whether participants remained more focused on price areas or discount areas. This analysis showed that all participants remained more focused on price areas. The participants that had a higher cognitive reflection spent more time on the price area compared with the discount area. This means a longer fixation on that specific area than on the discount in order to figure out the calculation. This also means a more in-depth analysis of the information because, as mentioned above, longer fixations are commonly associated with deep information processing.
As we said above, calculations need a deeper analysis of the information, so we expect that the Attentional Index predicts the accuracy of the choice. As a first analysis, we ran a mixed effects logistic model to examine if the Attentional Index is predictive of accuracy. We considered accuracy as a dependent variable and Attentional index (AIi) as an independent variable. The results showed that the Attentional Index was predictive of accuracy (Appendix C, Table 1) with a z-value = 2.070 and p-value = 0.038. With this analysis we found that this index predicted accuracy.

We confirmed the first hypothesis that the participants' analysis of the information, expressed by the Attentional Index, predicts the accuracy of the choice.

After this analyses we proceeded to analyze the results of the second hypothesis in order to understand if cognitive reflection, numeracy or both can predict the Attentional Index.

Hypothesis 2

In order to test the Hypothesis 2.1 and 2.2 we ran five different models (with all the combination among integral or limited versions of CRTs and integral or limited versions of numeracy). What we can see from the Table 2 (Appendix C), with regard to numeracy, is that only the integral version of numeracy (with the two questions from the Frederick's CRT test) is significant in order to predict the Attentional Index with a t-value = 1.817. We could explain this result giving the predictive power to the two Frederick's questions included into the integral version of numeracy. On the contrary both the CRT versions have a predictive power with a t-value = 2.622 for the integral version (with the three Frederick's items) and a t-value = 2.397 for the limited version (without the three Frederick's items). We decided to maintain the integral version of the CRT (with all the cognitive reflection tests because the Frederick's question was designed in order to measure this construct. A numeracy scale with also these questions would be a "dirty variable".)

In the model with both ("All CRTs with" and "Weller without") only the variable All CRTs is significant with a t value = 2.523.

Hypothesis 3

In order to confirm or disprove the third hypothesis we controlled four different points:

1. the Attentional Index predicts the accuracy (information obtained in order to test the first hypothesis Table 1, Appendix C; we reported this result also into the first model of the Table 3);

2. a regression analysis with CRT predicting Attentional Index where the relationship is significant (information obtained from the hypothesis 2.1;

model 3 in Table 2, Appendix C);
3. a regression analysis with CRT predicting accuracy where the relationship is significant (model 2 in the Table 3 Appendix C);
3. a multiple regression analysis with CRT and Attentional Index predicting accuracy (model 3 in Table 3, Appendix C).

We already have found that the Attentional Index predicts the accuracy (Hypothesis 1) and that the CRT predicts the Attentional Index (Hypothesis 2.1). In order to test the second and third point of the mediation analysis we computed a generalized linear mixed model where the accuracy is the dependent variable and the CRT (second point) or both CRT and the Attentional Index (third point) were the two explanatory variables. This last model was ran in order to test if the effect of the cognitive reflection on the accuracy could be explained by the Attentional Index.

As we can see from the second model of the Table 3 (Appendix C) the CRT predicted the accuracy with a z-value = 2.568 and a p-value = 0.0102.

In the third model of the Table 3 (Appendix C) with both variables, the CRT predicted the accuracy with a z-value = 2.238 and a p-value = 0.025 and the AI predicted the accuracy with a z-value = 1.682 and a p-value = 0.0925. The analysis shows that CRT is a more relevant predictor of accuracy, compared to the attentional index..

All the analyses mentioned above show that:

- the Attentional Index predicts accuracy.
- The CRT predicts the AI and numeracy (without the Frederick's questions) did not.
- In the model with both, CRT and AI to predict accuracy, the two variables remained significant; however, CRT is a more relevant predictor of accuracy, compared to the attentional index.

We decided to compute an extra analysis in order to answer another important research question that emerges from the analysis: "in a commercial problem-solving scenario involving high numerical components, does only numeracy affect the accuracy of the decisions?".

In the Table 4, Appendix C we presented the effect of numeracy (without the Frederick's questions), combined with the CRTs (with all the four versions) and the Attentional Index, on accuracy.

- Results of the models indicate that:
    - The limited version of numeracy measured by [8] predicts the accuracy with a z-value = 4.481 and a p-value = 0.0001;
    - In the model with the limited version of numeracy measured by [8] and the full version of CRT, numeracy has a significant predictive power (z-value = 3.602 and p-value = 0.0003) and all the CRTs taken together do not (z- value = 0.856 and p-value = 0.392).
    - In the model with numeracy and the Attentional Index, numeracy has a significant predictive power (z-value = 4.591 and p-value = 0.0001) and also the Attentional Index has a significant predictive power (z- value = 2.263 and p-value =0.0236).
    - In the model with all the variables only numeracy (z-value = 3.887 and p-value = 0.0001) and the Attentional Index (z-value = 2.124 and p-value = 0.0336) have a predictive power and the CRT has not.

These results indicated that the effect of numeracy is more relevant than the effect of CRT on accuracy. The role of numeracy in this particular economic decision-making process is predominant. However, the role of cognitive reflection on the attentional index is crucial and the role of numeracy on this variable is not significant. CRT is used as a proxy for the willingness and/or ability to engage in a more complex, effortful, and accurate strategy and this is reflected in their predictive power on the attentional index.

Four are the main findings of this research:

- In these tasks with high numerical components, the role of numeracy to predict the accuracy of the choice is predominant;
- The depth of analysis (measured by the Attentional Index) is crucial for the accuracy of the choice;
- The depth of analysis (measured by the Attentional Index) is independent from the numeracy;
- The depth of analysis is predicted by the cognitive reflection.

The last point highlight the role of cognitive reflection not only as an ability to inhibit a prepotent response, but also a measure of the propensity, inclination to a deeper and more reflective analysis of the information and reasoning.

Anyway, further research is needed in order to understand the link between cognitive reflection, attention allocation and numeracy.

5. Discussion

The main aim of this paper was to analyze how people deal with problems involving high numerical components when they have to find the most convenient option. Two scenarios were presented that were taken from daily life in order to understand the role of numeracy and cognitive reflection on accuracy. From the results we saw that not only numeracy, but also cognitive reflection played an important role in the accuracy of decision-making for these tasks. More specifically, cognitive reflection seems to play an important role in how we analyze information and on the level of attentional processing. People with lower cognitive reflection seem to analyze the pieces of information more superficially without going deeper; a cognitive impulsivity that leads to an approach closer to System 1, with fast and automatic intuition instead of System 2, which is more analytic and reflective [27]. Anyway the role of numeracy is crucial for the accuracy of the choice and numeracy is a more relevant predictor of accuracy, compared to the cognitive reflection. However, the crucial role of cognitive reflection on attention is still interesting and further research is needed to investigate whether and how CRT affects attention (and therefore accuracy) and the link between numeracy, strategy selection, and attention allocation in decision-making. Anyway, these analyses showed that cognitive reflection could be considered as an indicator of the level of engagement in more in-depth analysis. It is one thing is the ability to perform the computations leading to the accurate answer (numeracy) and another is the willingness to engage in more demanding reasoning (a part of cognitive reflection).

Certain groups of consumers are more vulnerable than others but the consumer's choice should be seen more as a resulting product of overlapping vulnerabilities and not as the result of only one force (e.g. age [28], self-control, gullibility, susceptibility [29], less education, marital status [30]). Vulnerabilities, as for example numeracy and cognitive reflection, could be seen as forces that act on the person, driving her performance. What we can see is only the resulting vector, the consumer in our case, with all her vulnerabilities. The freedom to choose, directly linked to the individual capability, reflects the level of an individual's resources, in the capability approach [31], [32]. What we argue with this paper is that the person's capability, and consequently what makes people vulnerable as consumers, comprises all the consumer characteristics and cognitive abilities (not only numeracy but also cognitive reflection and many others) that should be analytically researched in order to understand the resulting vector and the real power that gives freedom to that specific person. As a second step, a debiasing process is important in order to protect and to safeguard consumers and their needs during the consumption experience through consumer policy measures [28], [33]. Specifically, the attention of the European unfair commercial practices directive, the consumer protection law and consumer protection statutes increasingly recognize the importance of consumer vulnerability [34]. In recent years the European Parliament adopted several resolutions dealing specifically with consumer protection (e.g. resolution of 22 May 2012 on a strategy for strengthening the rights of vulnerable and resolution of 11 June 2013 on a new agenda for European Consumer Policy). The main point is to better understand the consumer vulnerabilities before targeting consumer policy measures. As we can easily understand, if we think about consumer vulnerability only as a numeracy issue we will develop a school support which deals, for example, only with arithmetic operations. Cognitive reflection, as we have seen from this study, plays an important role too and it should have a direct relevance for educators and policymakers.

Appendix A

Numeracy Scale

Taken from [8]:

(1) Suppose you have a close friend who has a lump in her breast and must have a mammography. The table below summarizes all of this information. Imagine that your friend tests positive (as if she had a tumor), what is the likelihood that she actually has a tumor?

(2) A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost? Cents ____

(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? Days ____

(4) In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?

(5) In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket from BIG BUCKS?

(6) Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up as an even number?

(7) If the chance of getting a disease is 20 out of 100, this would be the same as having a % chance of getting the disease.

(8) If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?

Appendix B

Cognitive Reflection Test

Taken from [10]:

(1) A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost? Cents____ *[Correct answer: 5 cents; intuitive answer: 10 cents]*

(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? Minutes____ *[Correct answer: 5 minutes; intuitive answer: 100 minutes]*

(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? Days____ *[Correct answer: 47 days; intuitive answer: 24 days]*

Taken from [12]:

(1) If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together? Days____ *[Correct answer: 4 days; intuitive answer: 9]*

(2) Jerry received both the 15^{th} highest and the15th the lowest mark in the class.

How many students are in the class? Students____ *[Correct answer: 29 students; intuitive answer: 30]*

(3) A man buys a pig for $60, sells it for $70, buys it back for $80, and sells it finally for $90. How much has he made? Dollars____ *[Correct answer: $20; intuitive answer: $10]*

(4) Simon decided to invest $8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has: a. broken even in the stock market, b. is ahead of where he

began, c. has lost money [*Correct answer: c; intuitive answer: b]*

Taken from [17]:

(1) If three elves can wrap three toys in hour, how many elves are needed to wrap six toys in 2 hours? *[Correct answer: 3 elves; intuitive answer: 6 elves]*

(2) Jerry received both the 15^{th} highest and the 15^{th} lowest mark in the class. How many students are there in the class? *[Correct answer: 29 students; intuitive answer: 30 students]*

(3) In an athletics team, tall members are three times more likely to win a medal than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes? *[Correct answer: 15 medals; intuitive answer: 20 medals]*

Taken from [18]:

(1) If you are running a race and you pass the person in second place are you in? *[Correct answer: second; intuitive answer: first]*

(2) A farmer had 15 sheep and all but 8 died. How many are left? *[Correct answer: 8; intuitive answer: 7]*

(3) Emily's father has three daughters. The first two are named April and May. What is the third daughter's name? *[Correct answer: Emily; intuitive answer: June]*

(4) How many cubic feet of dirt are there in a hole that is 3' deep x 3' wide x 3' long? *[Correct answer: none; intuitive answer: 27]*

Appendix C

Table 1: Accuracy as function of the Attentional Index

	Dependent variable:
	Accuracy
Attentional Index	0.302**
	(0.146)
Constant	0.729***
	(0.122)
Observations	3,518
Log Likelihood	−2,073.034
Akaike Inf. Crit.	4,154.069
Bayesian Inf. Crit.	4,178.731
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2: Attentional Index as function of numeracy and cognitive reflection

	Dependent variable:				
	Attentional Index				
	(1)	(2)	(3)	(4)	(5)
Weller with	0.274*				
	(0.151)				
Weller without		0.126			−0.088
		(0.167)			(0.179)
All CRTs			0.345***		0.380**
			(0.132)		(0.150)
All CRTs without				0.313**	
				(0.131)	
Constant	0.200**	0.286**	0.212***	0.230***	0.253**
	(0.097)	(0.113)	(0.066)	(0.065)	(0.108)
Observations	3,518	3,518	3,518	3,518	3,518
Log Likelihood	156.812	155.569	158.296	157.780	157.613
Akaike Inf. Crit.	−303.624	−301.138	−306.593	−305.561	−303.227
Bayesian Inf. Crit.	−272.796	−270.310	−275.764	−274.732	−266.233
Note:			*p<0.1; **p<0.05; ***p<0.01		

Table 3: Accuracy as function of the Attentional Index and cognitive reflection

	Dependent variable:		
	Accuracy		
	(1)	(2)	(3)
Attentional Index	0.302**		0.246*
	(0.146)		(0.146)
All CRTs		0.676**	0.585**
		(0.263)	(0.261)
Constant	0.729***	0.535***	0.486***
	(0.122)	(0.161)	(0.160)
Observations	3,518	3,518	3,518
Log Likelihood	−2,073.034	−2,071.991	−2,070.610
Akaike Inf. Crit.	4,154.069	4,151.983	4,151.221
Bayesian Inf. Crit.	4,178.731	4,176.646	4,182.049
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: Accuracy as function of numeracy, cognitive reflection and Attentional Index

	Dependent variable:			
	Accuracy			
	(1)	(2)	(3)	(4)
Weller without	1.251***	1.127***	1.215***	1.159***
	(0.279)	(0.313)	(0.265)	(0.298)
All CRTs		0.228		0.106
		(0.266)		(0.259)
Attentional Index			0.311**	0.299**
			(0.137)	(0.141)
Constant	0.025	0.003	−0.066	−0.073
	(0.209)	(0.209)	(0.204)	(0.204)
Observations	3,518	3,518	3,518	3,518
Log Likelihood	−2,066.557	−2,066.194	−2,064.109	−2,064.026
Akaike Inf. Crit.	4,141.114	4,142.389	4,138.218	4,140.053
Bayesian Inf. Crit.	4,165.776	4,173.217	4,169.046	4,177.046
Note:				*p<0.1; **p<0.05; ***p<0.01

References

[1] Stanovich KE, West RF. Individual differences in reasoning: Implications for the rationality debate? Behavioral and brain sciences. 2000;23(5):645–665.

[2] Kirsch IS, et al. Adult literacy in America: A first look at the results of the National Adult Literacy Survey. 1993.

[3] Kerka S. Not just a number: Critical numeracy for adults. ERIC Clearinghouse on Adult, Career, and Vocational Education, Center on Education and Training for Employment, College of Education, the Ohio State University; 1995.

[4] Fagerlin A, Zikmund-Fisher BJ, Ubel PA, et al. Measuring numeracy without a math test: development of the Subjective Numeracy Scale. Medical Decision Making. 2007; 27(5):672–680.

[5] Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Medical decision making. 2001;21(1):37–44.

[6] Schwartz LM, Woloshin S, Black WC, et al. The role of numeracy in understanding the benefit of screening mammography. Annals of internal medicine. 1997;127(11):966–972.

[7] Cokely ET, Galesic M, Schulz E, et al. Measuring risk literacy: The Berlin numeracy test. Judgment and Decision Making. 2012;7(1):25.

[8] Weller JA, Dieckmann NF, Tusler M, et al. Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. Journal of Behavioral Decision Making. 2013; 26(2):198–212.

[9] Peters E, Hibbard J, Slovic P, et al. Numeracy skill and the communication, comprehension, and use of risk-benefit information. Health Affairs. 2007;26(3):741–748.

[10] Frederick S. Cognitive reflection and decision making. The Journal of Economic Perspectives. 2005;19(4):25–42.

[11] Toplak ME, West RF, Stanovich KE. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. Memory &

Cognition. 2011;39(7):1275–1289.

[12] Toplak ME, West RF, Stanovich KE. Assessing miserly information processing: An expansion of the Cognitive Reflection Test. Thinking & Reasoning. 2014;20(2):147–168.

[13] Mata F, Miranda DM, Sallum I, et al. Do general intellectual functioning and socioe-conomic status account for performance on the Children's Gambling Task? Frontiers in neuroscience. 2013;7:68.

[14] Noori M. Cognitive reflection as a predictor of susceptibility to behavioral anomalies. Judgment and Decision Making. 2016;11(1):114.

[15] Graffeo M, Polonio L, Bonini N. Individual differences in competent consumer choice: the role of cognitive reflection and numeracy skills. Frontiers in psychology. 2015;6.

[16] Simonovic B, Stupple EJ, Gale M, et al. Stress and Risky Decision Making: Cognitive Reflection, Emotional Learning or Both. Journal of Behavioral Decision Making. 2016.

[17] Primi C, Morsanyi K, Chiesi F, et al. The development and testing of a new version of the cognitive reflection test applying item response theory. Journal of Behavioral Decision Making. 2015.

[18] Thomson KS, Oppenheimer DM. Investigating an alternate form of the cognitive reflection test. Judgment and Decision Making. 2016;11(1):99.

[19] Cohen J. A power primer. Psychological bulletin. 1992;112(1):155.

[20] Cokely ET, Kelley CM. Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. Judgment and Decision Making. 2009; 4(1):20.

[21] Finucane ML, Gullion CM. Developing a tool for measuring the decision-making competence of older adults. Psychology and aging. 2010;25(2):271.

[22] Velichkovsky BM. From levels of processing to stratification of cognition Converging evidence from three domains of research. Stratification in cognition and consciousness. 1999; 15:203.

[23] Velichkovsky BM. Heterarchy of cognition: The depths and the highs of a framework for memory research. Memory. 2002;10(5-6):405–419.

[24] Eivazi S, Bednarik R. Predicting problem-solving behavior and performance levels from visual attention data. In: Proc. Workshop on Eye Gaze in Intelligent Human Machine Interaction at IUI; 2011. p. 9–16.

[25] Glöckner A, Herbold AK. An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. Journal of Behavioral Decision Making. 2011;24(1):71–98.

[26] Bates D, Mächler M, Bolker B, et al. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:14065823. 2014.

[27] Sloman SA. The empirical case for two systems of reasoning. Psychological bulletin. 1996; 119(1):3.

[28] Berg L. Consumer vulnerability: are older people more vulnerable as consumers than others? International journal of consumer studies. 2015;39(4):284–293.

[29] Langenderfer J, Shimp TA. Consumer vulnerability to scams, swindles, and fraud: A new theory of visceral influences on persuasion. Psychology & Marketing. 2001;18(7):763–783.

[30] Lee J, Soberon-Ferrer H. Consumer vulnerability to fraud: Influencing factors. Journal of consumer affairs. 1997;31(1):70–89.

[31] Sen A. Commodities and capabilities. Lectures in economics: Theory, institutions. Policy. 1985;7.

[32] Sen AK. The idea of justice. Harvard University Press; 2009.

[33] Brennan C, Sourdin T, Williams J, et al. Consumer vulnerability and complaint handling: Challenges, opportunities and dispute system design. International Journal of Consumer Studies. 2017;41(6):638–646.

[34] Boom WH, Garde A, Akseli. The European Unfair Commercial Practices Directive: Impact, Enforcement Strategies and National Legal Systems. Routledge Taylor and Francis group; 2014.

STUDY 2

CHAPTER 3

Getting the best deal: The role of a topical mental accounting of multi-dimensional prices, and cognitive reflection

ABSTRACT

In this paper, we study the role of cognitive reflection on the mental accounting of choice attributes of an offered product/service package, and its impact on the accuracy of choice. Participants were presented with two websites selling the same package, and asked to choose one. One package had always a lower total price than the other. The two alternative packages were described by the same choice attributes. Some attributes were easily grouped together (similar attributes) whereas the other attribute was not (dissimilar attribute). Gaze behaviour and final choice were recorded. Last, participants were asked to fill several cognitive reflection tests. Results show that participants with lower cognitive reflection integrate less the dissimilar attribute and this is reflected in worse choice accuracy.

Keywords: *online purchases, eye tracking, cognitive reflection, categorization process, mental accounting.*

1. Introduction

"Mental accounting is the process, sometimes implicit, by which individuals and households keep track of and evaluate their transactions" ([1] p.12). The notion of mental accounting has spawned considerable research, both empirical and conceptual ([2] for a recent review).

A fundamental aspect of mental accounting is to describe how people mentally organize economic information: What gets combined with what. In the context of the evaluation of a purchase, for example, people tend to use a topical mental account of the advantages and disadvantages of the transaction ([3]; [4]; [5]). With that outcome-framing, only the outcomes related to the topic of the purchase are posted in the same mental account, and jointly evaluated. Several behavioural findings can be predicted from a topical organization of the elements of the decision problem.

In the classic calculator problem ([3]; [5]), only the prices (regular and discounted) associated to the topic of the purchase, the "calculator", are considered (e.g. the price of the other purchase, the jacket, is not considered). The use of the topical mental account rather than a comprehensive one (where the price of the jacket and other aspects such as current wealth or future earnings are considered) affects the likelihood to accept the same price discount. When people use a topical mental account rather than a minimal one of the same transaction (e.g. they only consider the difference in the two prices of the calculator) a similar effect is found. Another classic demonstration of mental accounting is the ticket problem ([3]; [5]). Here, people are more inclined to go to see a play when they lost a note than a pre-purchased ticket for the play of the same monetary value.

Henderson and Peterson ([6]) use the categorization theory to explain the typical findings in the calculator and the ticket problems as well as other choice anomalies due to mental accounting. They argue that (p. 97) "a mental account is merely a type of category that contains the gains and losses (advantages and disadvantages) of an event or element, similar to an evaluation category described by Fiske and Pavelchek ([7])". The principles that govern the grouping of elements in the categorization theory are the same principles guiding the inclusion of gains, losses, advantages and disadvantages into a mental account. For example, when the category "play" is mentioned people evoke that conceptual category which allows them to distinguish between conceptually relevant (e.g. the loss of a 10$ ticket) and irrelevant (e.g. the loss of a 10$ note) elements. This would also explain why, for example, in the ticket problem the loss of a 10$ note that was intended to be used to buy the ticket (a "ticket-directed note") is more easily posted to the same mental account of the purchase of the ticket than the loss of a note of the same monetary value ([8]).

People spontaneously use conceptual similarity cues to group things such as expenses, words, risks, gains or losses (see [9] for a review on the grouping and labelling of resources). For example, consumers spontaneously group expenses into separate mental accounts (e.g. “clothing”, “food and entertainment”. See [10]). Anecdotes show that households often structure their shopping list by organizing planned purchases by a conceptual similarity criterion: ‘Bananas’ are written on the list near to ‘apples’, and not near ‘electric plugs’.

People make distinctions in terms of types of wealth such as “current spendable income”, “current assets” and “future income” ([11]). They also categorize types of saving/gain, and this categorization affects their future choices. For example, Henderson and Peterson ([6]) found that money saved from one category (e.g. cash refund for returning a record album) was more likely to be spent on the same category than a different category.

People also seem to group losses in terms of their conceptual similarity. Singer et al. ([8]) findings suggest that the loss of a “ticket-directed note” is more similar to the loss of a “ticket” than the loss of a “note”. Because of their semantic associative strength, the first two losses are more easily grouped together (and posted to the same mental account) than the third loss.

Last, people group semantically related prices, and risky outcomes. By using different versions of the calculator problem, Bonini and Rumiati ([12]) shown how the use of a comprehensive mental account (e.g. the integration of the prices of the two purchases) is contingent upon the similarity structure of the problem. A comprehensive mental account is used when the two planned purchases belong to the same basic category level (e.g. a classic and casual shirts) and not when they relate to different categories (e.g. a classic shirt and a camera). A similar finding is reported in Bonini, Tentori and Rumiati ([13]) where people combine risky outcomes of a choice option only when the outcomes are semantically related (e.g. contraction of a virus alpha with probability p and contraction of a virus beta with probability q).

In terms of basic cognitive processing, the grouping of elements by their semantic relatedness is congruent with the spreading activation model of semantic memory. This model predicts that the activation of a conceptual category first spreads to strongly related concepts ([14]). Meyer and Schvaneveldt ([15]) have found evidence of a semantic priming effect.

Also, conceptual categories are used to spontaneously organize the free recall of a list of randomly presented words that have to be memorized ([16]). As argued by Tulving ([17], p. 352), the spontaneously subjective organization of words in clusters (e.g. words related to category "Fruit" and words related to the category "Cloth") help people to accurately memorize them.

In sum, there is ample experimental evidence and theoretical elaboration on the use of conceptual similarity to group the elements of a decision problem (or to organize a memory task). In this paper, we study how people code multi-dimensional prices of alternative commercial offers in the presence of a conceptual similarity structure, and how this editing activity affects the quality of their choice (e.g. the selection of the best offer). Consistent with the previously discussed literature, we argue that people spontaneously group semantically related prices of a transaction before making the final choice, and this type of coding will affect the quality of their commercial choice (conceptual similarity hypothesis).

To clarify the object of our study, and our research predictions consider the following commercial offer presented by a multi-dimensional price (Hooman Estelami [18], [19], quoted in [20]). If you accept it, you get a brand new car at the price of 20,000€, with the automated-opened roof at the price of 4,000€, with the special metallic paintwork for 1,000€, and a loan which entirely finances the purchase of the car at the 4 percent rate. Take it or leave it: No room for bargaining. This is a bundle offer, and is not negotiable.

If people use a "comprehensive mental account" of the commercial offer, they might incorporate several aspects such as current wealth, future earnings, and integrate all price information of the transaction, including the financial cost for the buying of the car. [footnote 1: In this experiment, we are not studying the hedonism editing principles of mental accounting ([21]). For example, we are not comparing a partitioned price with the corresponding all-inclusive price. However, if price/payment information is considered a proxy for a loss (e.g. they all entail a displeasure) then people should consider the all-inclusive price of the transaction, and not its separate price components (or a subset of) because people prefer integrated than segregated losses (evidence supporting that prediction in the domain of bundle prices is reported in Johnson, Herrmann and Bauer ([22]), ([20]). However, see Abraham and Hamilton, ([23]) for a recent meta-analysis where mixed findings are reported]. This is not a negligible cost. In fact, it amounts to more than 10% of the true price of the car, that is 2,600€ in interest over its life.

People might differently structure the decision problem by an alternative coding of the multi-dimensional price. They might group semantically related prices, and segregate the unrelated ones. For example, they integrate the first three prices because they all relate to the category "car": the price for the car, and the price for its special physical features. With this coding, they segregate the loan cost because it is perceived as not semantically related to the topical object "car" (e.g. loans are something related to finance and banks, not to cars). This mental representation of the prices of a transaction is coherent with the notion of "topical mental account", and with a more general interpretation of mental accounting process as category-based. Said differently, the grouping of prices is based on their categorical relevance (e.g. whether or not they relate to the topic or category of purchase).

Evidence from the car loan market shows that many consumers do not take in consideration how the cost of the loan will affect the final cost of the purchase. Most people overpay for car loans and some car companies received a fine for their commercial policy; as described by the president of the National Consumer Union, some car companies tried to promote car sales in conjunction with a high interest loan to the consumers ([24]). Most cars are purchased with the help of a car loan and most consumers are taking out car loans at interest rates that are much higher than they could have gotten if they had just shopped around more ([25]), and this happens also in the house market ([26]). Consumers take the first interest rate that is offered to them; they did not integrate the cost of financing the car purchase in the total car cost, making a disadvantageous purchase. There might be several reasons why people spend so much time and effort to look for the best-featured car, and dedicate so little time for searching the best loan. One of them is that consumers are focused on the object "car" and its mentally inherent features ("topical mental account"), and disregard aspects that are perceived as not directly related to the car itself (e.g. the type of financial plan to buy it, future earnings, etc.).

Problem structuring matters. If people do not integrate the semantically unrelated prices, they may be tempted to accept the transaction that looks very good on the semantically-grouped prices, although it is very bad on the other prices. Moreover, when asked to make a choice between two offers X and Y where the first is better than the latter on the grouped- prices but worse totally (e.g. X is more expensive than Y when all prices are considered), people might be tempted to choose the worst commercial offer.

As said, the aim of this paper is to investigate how people code a multi-dimensional price. Specifically, we want to control whether people segregate semantically unrelated prices of a commercial offer, and how this affects the quality of their choices. Our main research hypothesis is that the presence of a semantic relatedness structure in the decision problem affects both the coding of prices, and the resulting quality of the choice. Specifically, we make the hypothesis that (i) when a semantically unrelated price is displayed, people segregate it from the other semantically related prices of the transaction. Also, (ii) the final choice is a function of the relative cost of the semantically- grouped prices. For example, if the cost of the grouped prices for X is lower than for Y, then people will choose X -even when X is more expensive than Y. Finally, (iii) the two predicted findings (i) and (ii) are expected to be more pronounced for impulsive than reflective consumers. This third prediction is based on the fact that the scientific literature has consistently found that cognitive reflection correlates with heuristics and biases ([27]; [28]; [29]); people with low CRT scores base their judgments and decisions on System 1 processes (fast, automatic, effortless and related to intuitive thoughts). Since highly impulsive people tend to be overridden by System 1 processes (fast, automatic and effortless), which ultimately make them more vulnerable in many ways ([30]), we expect that they will be more affected by similarity basic assessments (considered to be a typical System 1 process, see [31] for a discussion) even when the latter are not informative and useful in order to "solve the problem" (e.g. getting the best deal).

The aim of this paper is to explore the role of similarity perception in the domain of economic choice. If most of the studies on mental integration vs. segregation focus on the mental accounting of different products or services, here we study whether similarity perception affects the mental accounting of different attributes within a given product or service. We argue that attributes that are typically related to a good (e.g. its price or qualities) are more easily grouped together than attributes that are considered to be external to the same good (e.g. an extra warranty). For example, people might be presented with a commercial offer for a printer bundle where they are provided with the following cost components: the printer cost, the transport cost, the installation cost, and the extended warranty cost. We argue that the first three cost components are more easily grouped together (e.g. the costs for "printer") than the fourth cost (e.g. something "external to the printer"). If people rely on this intra-product mental accounting then they will tend to evaluate the deal based on the spontaneous grouping of the two types of cost components: the similar vs. the dissimilar ones. This similarity based mental accounting of product attributes will ultimately affect choices (e.g. the distribution of the cost of attributes between the similar vs. dissimilar cost groups can be manipulated). If people rely on this mental accounting of the cost of attributes for a given deal, then they might overlook the fact that the cost for the dissimilar attribute is too high.

In this paper we want to control whether there are differences between high vs. low cognitive reflection individuals in the mental accounting of choice attributes for the same package deal, and if this ultimately affects the accuracy of their choices. Mental accounting processes will be assessed through the analysis of gaze behaviour whereas the accuracy of choice will be measured by the proportion of optimal choices (i.e. the choice of the product/service package with the less total final price).

2. Methods

2.1. Pilot study

The aim of the study was to understand how people categorize choice attributes of an offered product/service package (e.g., a printer package). Specifically, we wanted to detect those choice attributes that are easily grouped together compared to those that are not. To achieve this goal, a card sorting task was performed by the participants to the pilot study. The card-sorting method has been developed for designing mainframe menu systems ([32]), and has been recently used in the user centered design field in order to create information architectures of effective and user-friendly websites ([33]; [34]). We collected data as suggested in Tullis and Wood ([35]). For each commercial package, participants were presented with six set of cards (each card representing a choice attribute), and asked to group the cards by following a criterion which made sense for them. The names of the choice attributes to be grouped were printed on individual cards in an easily readable font. In Table 1 are shown the choice attributes used for each commercial package.

Washing machine	Gym membership	Holiday by the sea
room charge	WM cost	GM cost
tourist tax	transport cost	registration card cost
breakfast	mounting cost	discount
credit card fee	extended warranty	sauna
discount	disposal cost	locker cost
wi-fi fee	appropriate detergent	doctor’s appointment

Table 1. Choice attributes used for each commercial package.

To detect which choice attributes are easily grouped together and which are not, we computed how many times each attribute was grouped together with all the others. Results are reported in Table 2, 3 and 4 (Appendix A). In order to visualize the results, we used Cluster Dendograms (Figure 1, Figure 2, Figure 3) that represent the hierarchical cluster analysis ([36]).

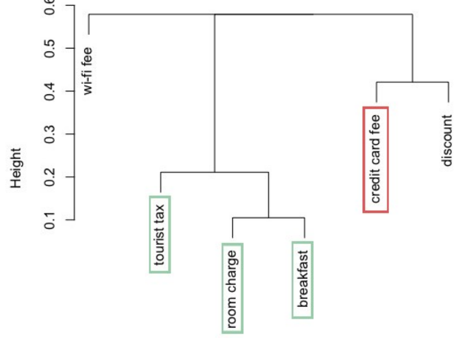

Figure 1. Cluster Dendogram for the "holiday by the sea" task.

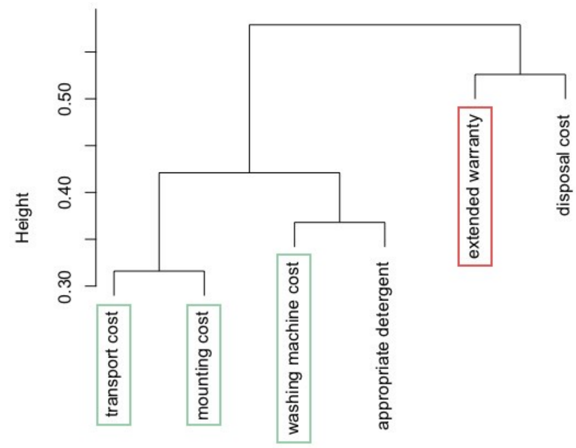

Figure 2. Cluster Dendogram for the "washing machine" task.

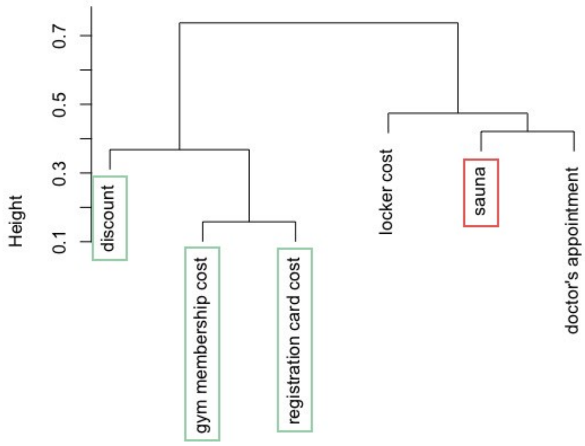

Figure 3. Cluster Dendogram for the "gym membership" task.

We chose as similar attributes (in green) the three closest ones, and as dissimilar attribute (in red) the one farthest from the similar ones (based also from the frequency tables in the Appendix A; in the first column are reported the four choice attributes for each commercial package used in the next Experiment).

On the basis of these results, in the next experiment each package was described by four choice attributes, that is the three similar attributes and the fourth dissimilar one. The three similar attributes were those that were more often grouped together. The dissimilar attribute was that which was less often grouped with the three similar ones. Let us call the first three attributes (p, q and r) "similar choice attributes", and the last attribute (s) "dissimilar choice attribute". Table 2 (Appendix A) summarizes the four choice attributes for each commercial package used in the next Experiment.

2.2. Experiment

2.2.1. Participants and procedure

The subjects who participated in the study were 54 university students (34 female, 20 male; mean age= 23.6 years, SD=4.5). We collected as many participants as our resources allowed to increase statistical power. The decision-making task was made incentive-compatible; during the instructions we told them that they will receive 3 euros for participation and from a minimum of 3 euros to a maximum of 11,5 euros on the basis of their performance. Every participant, after reading the instructions, took part in a decision-making section and afterwards, took four different CRTs without any time limit. We recorded participant eye movements during the decision-making part of the experiment, while eye movements during the cognitive reflection test part were not recorded. Eye movement data were recorded using an Eye Link 1000 Plus Binocular Tower Mount, which provides a data acquisition at up to 2000 Hz. The participants completed the decision-making task in the psychtoolbox interface on a Dell computer with a 23 inches screen.

2.2.2. Decision-making task: choosing where to purchase

After reading the instructions, participants were shown a product/service package, which was available in two different online shops. The two websites were offering exactly the same product/service but with a different distribution of costs across the same four choice attributes. The two alternative shops were presented in the rows, and the four choice attributes were presented in the columns. For each item, one of the websites had a better promotion than the other (i.e. the full total cost of the package was less). Specifically, 50% of the time the "X website" was the best option, and

50% of the time the "Y website" was the best.

In Figure 4, the decision scenario related to the printer's package (which was used as a training choice trial) is shown. Choosing between bundled commercial packages is quite common in the online market as well as in the utility market. For example, an offer is valid only if you choose the same company for both gas and electric power.

Participants were presented with three types of decision scenarios. Each one was related to a different commercial package: holiday by the sea, washing machine and gym membership. We considered as a between subject factor, the position of the dissimilar choice attribute "s" in the matrix. This was done in order to control a possible order effect. If we had left the dissimilar attribute always in the last column, the not integration effect could be explained by an effect order. Therefore, in the first condition all the participants saw the "s" element in the fourth column; in the second condition all the participants saw the "s" element in the first column, and in the third condition all the participants saw the "s" element in the third column.

Printer				
	Printer's price	Transport	Installation	Extended warranty
X website	580 €	15 €	5%	55 €
Y website	460 €	30 €	30 %	52 €

Figure 4. Decision-making part: Printer scenario explained during the instructions.

Since each subject was assigned to one and only one of the three conditions, the position of "s" is a between subjects factor; so the subjects had been divided in three groups, or conditions, depending on the position they would see the element "s". We decided to consider the position of the dissimilar attribute as a between subject factor in order to avoid a possible disorientation effect on the participants.

Leaving all the attributes value as numbers (not in percentage), the computation would be too easy and most likely all the attributes would be integrated between them. So, for each package, we built three different variants where one or two component costs was/were expressed in percentage terms. Specifically, in one variant, the cost of the choice attribute "s" was in percentage, in the second variant the cost of choice attribute "r" was in percentage, and in the third variant both choice attributes "s" and "r" were expressed in percentage terms. This was constant for all three decision scenarios, and was done in order to control for a possible not integration effect due to the fact that the "s" element was expressed in percentage. Each participant was presented with these 3 versions of the 3 packages for a total of 9 decision scenarios. Each participant saw 3 different scenarios and for each scenario 3 different variants depending on which attribute was expressed in percentage (r, s or both); depending on the condition, in all the 9 decisions, she saw the "s" attribute in the same column.

As far as the choice of the website is concerned, 50% of the time the "X website" was the best option, with the best package deal, and 50% of the time it was not. 50% of the time integrating the "s" attribute to "p", "q" and "r" was necessary in order to choose the best option, 50% of the time such integration made no difference. Let us call the tasks in which the "s" element was the swing attribute "crucial tasks". For each participant, four out of nine tasks were "crucial tasks".

2.2.3. *Cognitive Reflection Tests*

We used four different Cognitive Reflection Test (CRT) in order to have a stronger measure of this construct and to address the problem of the diffusion of the answers to the original CRT by Frederick (2005) [37]. Participants completed Frederick's test, Toplak et al., (2014) [38], Primi et al. (2015) [39] and Thomson and Oppenheimer (2016) [40]. Cognitive reflection was scored as the proportion of the total items answered correctly.

2.2.4. *Research Hypotheses*

The aim of the study was to determine whether a choice attribute (“s”) that is perceived dissimilar from the other three attributes ("p", "q" and "r") for a given commercial package, is integrated or kept separated during a choice process (mental accounting of choice attributes). As discussed in the Introduction, low CRT people rely heavily on automatic and intuitive processes (e.g. similarity basic assessments). Thus, we expect that:

H_1: Low CRT people tend to mentally segregate the fourth dissimilar choice attribute from the three similar ones.

Segregation of the dissimilar choice attribute will be measured by the average dwell time spent in the "s" attribute and the number of fixations in that attribute. We expect that, compared to high CRT people, low CRT people will fixate less often and less frequently on the dissimilar choice attribute “s” while deciding which package to buy.

H_2: As a consequence of failures to integrate the dissimilar attribute with the similar ones, the choice accuracy of low CRT people will be lower than high CRT people.

This is due to the fact that low CRT people tend to consider a partial total cost of the deal (the sum of the costs of the three similar choice attributes) instead of the full total cost (the sum of the four cost components) when making a choice (a choice based on a mental accounting). Choice accuracy is measured as the choice proportion of the better deal for each presented commercial offer (e.g. the package with the lowest final total price).

3. Results

3.1. CRT and eye movements

As measures of integration of the "s" element, we utilized the average dwell time in "s" and the number of fixations in "s". We choose to use the average dwell time instead of the total dwell because the total dwell time and the number of fixations on "s" are by nature correlated. Figure 5 shows the scatter plot of the average dwell time and of the number of fixations on "s", on the original (left) and log scale (right). The vertical and horizontal lines indicate the median values of average dwell time and number of fixations. We display the data also in the log-scale, because it is easier to discern the different data points in proximity of the median point. The data relative to the high CRT subjects (upper quartile) are in orange, and the ones relative to the low CRT subjects (lower quartile) in green. We perform correlation tests (Spearman's for the values on the original scale [r=0.025, p = 0.579] and Pearson's correlation on the log-values [r=0.009, p = 0.839]) and conclude that we cannot reject the hypothesis of zero correlation. As one can qualitatively grasp from Figure 5, there seems to be a relationship between CRT and both the average dwell time and the number of fixations. In particular, the majority of the data lying in the first quadrant (high average dwell time, high number of fixations) belong to high CRT individuals.

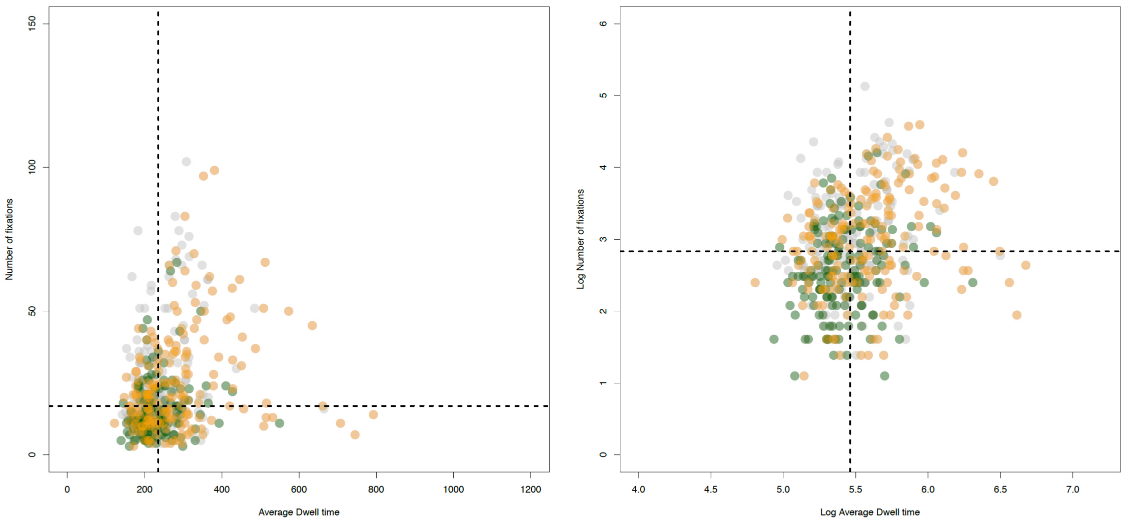

Figure 5. Scatter plot of average dwell time vs number of fixations on "s", on the original (left) and log scale (right). The data for HCRT subjects are in orange, for LCRT subjects in green.

As a representation of the average dwell time in the group of high CRT and low CRT we reported the heat maps of high CRT (Figure 6) and low CRT (Figure 7), representing the average dwell time of the two groups in the same trial (7^{th} trial) and in the same condition (condition one; the attribute "s" is in the last column).

As we can see from these figures, the average dwell time spent in "s" in the group of high CRT participants is higher compared to the average dwell time of the low CRT participants, in the same trial.

Therefore, we decided to quantitatively investigate the effect of CRT on the average dwell time and on the number of fixations in "s". Here, we would like to point out that, while Figure 5, Figure 6 and Figure 7 distinguish only between high CRT and low CRT subjects, in our quantitative analysis we consider CRT as a metric variable, including for each subjects the specific value of CRT.

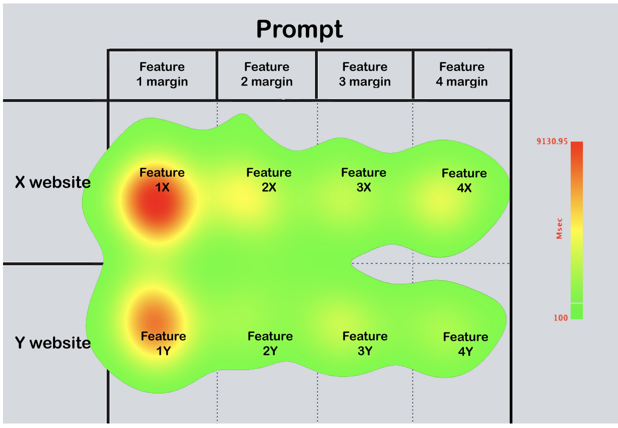

Figure 6. Heat map of the participants with high CRT in the 7^{th} trial.

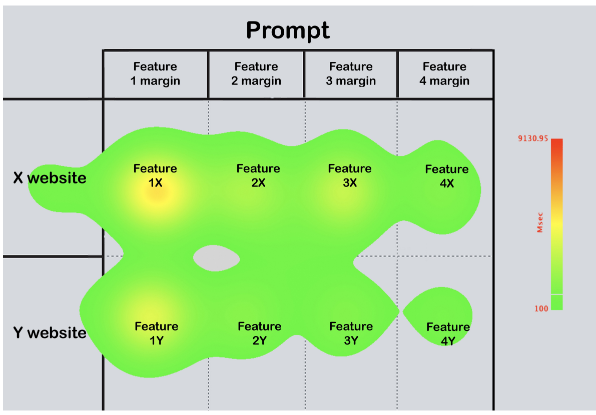

Figure 7. Heat map of the participants with low CRT in the 7^{th} trial.

3.1.1. CRT and average dwell time

To study the dependence of the average dwell time on the CRT, we propose a linear mixed effect model with a dependent variable of logged average dwell time and an independent variable of CRT value. We used the natural logarithm transformation of the average dwell times values in order to approximate a normal distribution [41]. Moreover, to account for the within-subject and within-trial correlation structure, we use additive non-nested random effects, respectively subject and trial specific. Finally, the between-subject condition is incorporated in the linear mixed effect model fixed effects, because we selected three conditions out of four possible conditions. We control for the condition by adding it as predictor on the right-hand side of the model equations. In Table 1 (Appendix B) the outputs of three different mixed linear regression models are reported. In the first model, we control for the between-subjects condition by modelling it through two dummy variables and their interaction with CRT, hence having condition specific intercepts and slopes; in the second model, we allow only the intercept to be condition specific; in the third model, we omit the control on the condition. According to the AIC and BIC criteria, the best model is the one in which the between subject condition fixed effect is omitted. After we have appropriately modelled the variance structure through the random effects, and we ascertained that the subject specific coefficient condition can be omitted from the model, we now focused on the effect of CRT on the log average dwell time in "s". By looking at the third model, it appears clear that there is a significant effect of CRT on average dwell time ($\hat{\beta} = 0.486, t = 3.72, p < 0.01$). In particular, for an increase in CRT of ΔCRT we expect a relative change in the average dwell time of $100(e^{0.486\Delta CRT} - 1)\%$. This corresponds to an expected relative change in the average dwell time of 3.53%, for each additional right answer in the CRT test.

3.1.2. CRT and number of fixations

To study the dependence of the number of fixation numbers on the CRT, we use a generalized mixed effect model, specifically a Poisson model with a logarithmic link function. The count dependent variable is the number of fixations, and the independent variable of interest is the CRT value. As done previously, we account for the within subject and within trial correlation structure through additive non nested random effects, respectively subject and trial specific, and we control for the between subject condition. In Table 2 (Appendix B) the outputs from three different mixed effect Poisson regression models are reported. As before, in the first model, we control for the between subjects condition by modelling it through two dummy variables and their interaction with CRT; in the second model, we allow only the intercept to be condition specific; in the third model, we omit the control on the condition. The significance of the parameters shows support for a significant difference in the intercept for the second condition and the first condition. However, on the basis of the model selection criteria, and following the principle of parsimony, there is support for choosing the third model. The third model shows a significant effect of the CRT on the number of fixations ($\hat{\beta} = 0.614, z = 1.986, p < 0.05$). When the CRT increases by ΔCRT, the predicted relative increase in the fixation count increases by $100(e^{0.614 \Delta CRT} - 1)\%$. This means that, when the number of corrected answers in the CRT increases by one, we expect a relative change in the number of fixations of 4.48%.

In this quantitative analysis we consider CRT as a metric variable, but we reported in Figure 8 and Figure 9 the transitions of a high and low CRT participant in the 7th trial in the first condition (the attribute "s" is in the last column) in order to visualize the results.

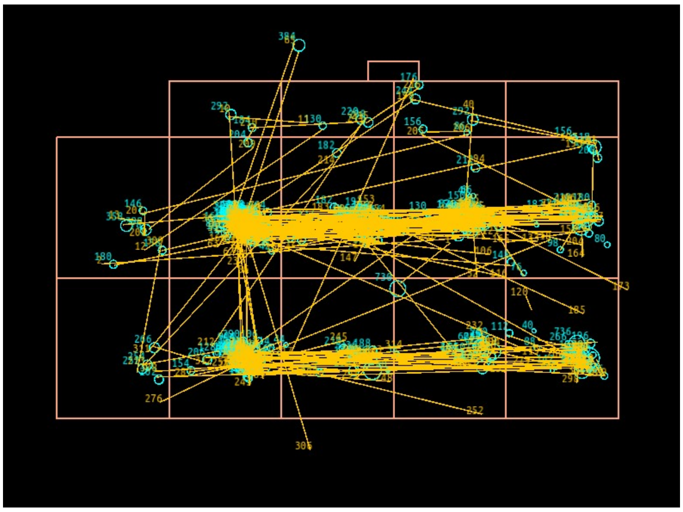

Figure 8. Fixations and saccades of a high CRT participant in the 7th trial.

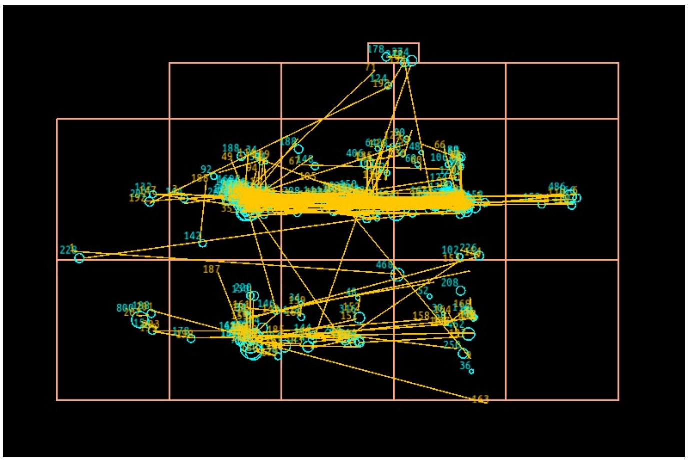

Figure 9. Fixations and saccades of a low CRT participant in the 7th trial.

As we can see, the number of transition in "s" increases increasing the level of CRT.

3.2. Choice Accuracy

3.2.1. Choice Accuracy and CRT

We study the relationship between choice accuracy and CRT by means of generalized mixed effect model, in particular we perform a mixed effect logistic regression. The response variable is binary and indicates if the trial for a specific subject was answered correctly. As previously done, we control for the within effect of subject and trial through additive random effects, and we introduce the between conditions through dummy variables. Moreover, we add a dummy variable, which indicates whether the trial was critical. A trial was defined as critical if the role of the "s" was crucial for the choice; more deeply, if the integration of the value of "s" leads to switch the website preference. In Table 3 (Appendix B) the outputs from three different mixed effect logistic regression models are reported. The sixth model is the one with lower AIC and BIC. The effect of the CRT on the choice accuracy is significant ($\hat{\beta} = 1.486, z = 3.342, p < 0.001$). The coefficient relative to the CRT in this model indicates the expected log odds of the probability of making the right choice, when the CRT goes from 0 to 1. This means that considering people with two levels of CRT which differ of ΔCRT, brings to an expected odds ratio $e^{1.486 \Delta CRT}$. Answering correctly to one more question on the CRT gives an estimated odd ratio of the probability of making the right choice of 1.111.

3.2.2 Additional Analyses

We decided to compute additional analyses in order to deal with a possible interpretation of the concept of “similarity”. The analysis mentioned above are based on a semantic similarity. The “s” element is different to the others because it was classified as not belonging to the same category of “p”, “q” and “r” from a semantic point of view. Anyway, a numerical dissimilarity could also emerge from the comparison between these elements and an important question to answer could be: “Does numerical format dissimilarity matter?. An element could be considered as different to the others because it is in percentage and the others are not. We decided to analyze this aspect in order to understand if the numeric format dissimilarity, the within subject factor, leads to different attentional patterns. In the first condition only the “r” element is in percentage; in the second condition (perc2 in the tables) only the “s” element is in percentage. In the third condition (perc3 in the tables) both, “r” and “s” elements are in percentage. We have also taken into consideration that this within subject factor (the element in percentage) might interact with the individual numerical abilities. So we decided to analyze the numeracy score measured by the 8-item numeracy scale developed by Weller et al. ([42]) and its interaction with the numerical format of the elements. We analyzed also the position of the "s", between subject factor, in order to check a possible order effect. In the first condition, the "s" element is in the last column; in the second condition (pos2 in the tables) the "s" element is in the first column and the condition three (pos3 in the tables), the "s" element is in the third column. As we can see from the tables (Table 1, 2 and 3, Appendix C):

- the percentage format has a greater effect on the number of fixations rather than dwell time;
- the interaction between the percentage format and numeracy is more significant compared to the interaction between the percentage format and the CRT score;
- the fixation length depends more on numeracy (or rather on the interaction between numeracy and percentage format);
- In terms of accuracy, both CRT and numeracy are good predictors.

3.2.2. Choice Accuracy and Eye Movements

We once again used a mixed effect logistic regression to analyze the relationship between choice accuracy and eye movements. Here, the independent variables of interest are the average dwell time and the number of fixations in "s". To avoid convergence problems of the numerical solver, we re scaled the independent variables to be between 0 and 1. This does not involve any loss of information, and simply needs to be accounted for in the interpretation of the parameters. We ran a series of logistic regressions, whose outputs can see in Table 4 (Appendix B). We now focus on the fourth model in Table 4. In the fourth model, we see that the significant independent variables are the number of fixations ($\hat{\beta}_{FIX} = -4.959, z = -2.500, p < 0.05$), and the interactions of the number of fixations and the average dwell time ($\hat{\beta}_{INT} = 27.8740, z = 2.500, p < 0.01$). The fifth and sixth model only account for the effects of number of fixations and the average dwell time, respectively. What is interesting is that while the number of fixations had a significant effect in the model with the interaction, it does not in the model in which it appears alone. On the other hand, the effect of the average dwell time which, conditioned on the presence of the number of fixations in the model, was significant only in the interaction, is now significant per se. The seventh model uses only the product as independent variable, but does not perform better than the model with only the average time as independent variable. The best model overall remains the fourth model, in which we consider the number of fixations and the interaction between the number of fixations and the average dwell time. It is not straightforward to interpret the coefficients of this model, since we do have a term, which is not linear in the parameters, and hence the effect of one independent variable will depend on the level of the other one. As seen before, in the context of logistic regression, one can think in terms of odds ratios, i.e. the chances of answering correctly before and after the increase in the independent variable. When the number of fixations increases by ΔFIX , we expect an odds ratio of $e^{-4.959\Delta FIX + 27.874\Delta FIX\times DWELL}$, which depends on the value of the average dwell time. This means that the strength of the effect of the number of fixations depends on how long they are on average. The same goes for the interpretation of the effect of the average dwell time. Increasing the average dwell time of $\Delta DWELL$ gives an estimated odds-ratio of $e^{27.874FIX\times\Delta DWELL}$.

In additional analyses (Table 4, Appendix C) we combined cognitive reflection with both eye movements measures, number of fixations and the average dwell time.

4. Discussion

Nowadays, an ever-increasing proportion of consumers makes online purchases where they are constantly inundated with information, instant communications and unfiltered messages. The selection of pertinent information and their analysis require particular abilities during the purchasing process. In the online purchase market there are offers that require a consumer to purchase bundled services or products. In that light we analyzed if people with a higher cognitive impulsivity are more inclined to integrate only attributes belonging to the same category when calculating the final price to pay. The results have shown that people with lower cognitive reflection less often integrate (segregate) the attribute that does not belong to the same category of the others. This means they perform an incomplete calculation that leads them to choose the less advantageous option (when that attribute matters for the amount of the final price).

Therefore, from this study, we highlighted that cognitive reflection influences the mental accounting process and the consequent estimate and choice of the best deal. Certain groups of consumers are more vulnerable than others. The European unfair commercial practices directive, the consumer protection law and the consumer protection statutes recognize the importance of consumer vulnerability, and they have implemented consumer policy measures in order to protect and safeguard consumers ([43]). From this perspective, a deeper analysis of the individual characteristics is advised in order to have a more complete view of the forces that act on the freedom to choose. Cognitive reflection, as we have seen from this study, plays an important role during the analysis and decision-making process, and it should have a direct relevance for educators and policymakers.

Appendix A

Typology		room charge	tourist tax	breakfast	credit card fee	discount	wi-fi fee
similar attribute	**room charge**	**1**	**0.789**	**0.895**	**0.053**	0.368	0.316
similar attribute	**tourist tax**	**0.789**	**1**	**0.789**	**0.211**	0.368	0.316
similar attribute	**breakfast**	**0.895**	**0.789**	**1**	**0.105**	0.421	0.421
dissimilar attribute	**credit card fee**	**0.053**	**0.211**	**0.105**	**1**	0.579	0.316
rejected attribute	discount	0.368	0.368	0.421	0.579	1	0.263
rejected attribute	wi-fi fee	0.316	0.316	0.421	0.316	0.263	1

Table 2. Pilot study results for the vacation task.

Typology		washing machine cost	transport cost	mounting cost	extended warranty	disposal cost	appropriate detergent
similar attribute	**washing machine cost**	**1**	**0.474**	**0.579**	**0.105**	0.316	0.632
similar attribute	**transport cost**	**0.474**	**1**	**0.684**	**0.421**	0.316	0.368
similar attribute	**mounting cost**	**0.579**	**0.684**	**1**	**0.368**	0.263	0.474
dissimilar attribute	**extended warranty**	**0.105**	**0.421**	**0.368**	**1**	0.474	0.316
rejected attribute	disposal cost	0.316	0.316	0.263	0.474	1	0.211
rejected attribute	appropriate detergent	0.632	0.368	0.474	0.316	0.211	1

Table 3. Pilot study results for the washing machine task.

Typology		gym membership cost	registration card cost	discount	sauna	locker cost	doctor's appointment
similar attribute	**gym membership cost**	**1**	**0.842**	**0.632**	**0**	0.105	0.211
similar attribute	**registration card cost**	**0.842**	**1.000**	**0.579**	**0.053**	0.211	0.263
similar attribute	**discount**	**0.632**	**0.579**	**1.000**	**0.158**	0.211	0.158
dissimilar attribute	**sauna**	**0**	**0.053**	**0.158**	**1.000**	0.526	0.579
rejected attribute	locker cost	0.1052632	0.211	0.211	0.526	1.000	0.263
rejected attribute	doctor's appointment	0.2105263	0.263	0.158	0.579	0.263	1

Table 4. Pilot study results for the gym membership task.

Appendix B

	Dependent variable:		
	log(av_time_fix)		
	(1)	(2)	(3)
CRT	0.616***	0.472***	0.486***
	(0.217)	(0.134)	(0.131)
Condition 2	0.042	−0.011	
	(0.170)	(0.077)	
Condition 3	0.355	0.044	
	(0.227)	(0.080)	
CRT:Condition 2	−0.095		
	(0.298)		
CRT:Condition 3	−0.548		
	(0.377)		
Constant	4.997***	5.073***	5.075***
	(0.130)	(0.094)	(0.081)
Observations	485	485	485
Log Likelihood	−223.589	−224.255	−221.119
Akaike Inf. Crit.	465.178	462.511	452.237
Bayesian Inf. Crit.	502.835	491.800	473.158
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 5. CRT and average dwell time.

	Dependent variable:		
	Count_Fix		
	(1)	(2)	(3)
CRT	1.206**	0.623**	0.614**
	(0.482)	(0.303)	(0.309)
Condition 2	0.792**	0.302*	
	(0.377)	(0.172)	
Condition 3	0.782	0.282	
	(0.502)	(0.180)	
CRT:Condition 2	−0.953		
	(0.661)		
CRT:Condition 3	−0.910		
	(0.834)		
Constant	2.307***	2.617***	2.812***
	(0.287)	(0.208)	(0.187)
Observations	485	485	485
Log Likelihood	−2,575.565	−2,576.725	−2,578.556
Akaike Inf. Crit.	5,167.131	5,165.450	5,165.111
Bayesian Inf. Crit.	5,200.604	5,190.555	5,181.848
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6. CRT and number of fixations.

	Dependent variable:					
	Accuracy					
	(1)	(2)	(3)	(4)	(5)	(6)
CRT	2.079**	1.688**	1.688**	1.537***	1.537***	1.486***
	(0.823)	(0.719)	(0.719)	(0.439)	(0.439)	(0.445)
Condition 2	−0.256	−0.258	−0.258	−0.300	−0.300	
	(0.538)	(0.538)	(0.538)	(0.250)	(0.250)	
Condition 3	−0.146	−0.145	−0.145	−0.481*	−0.481*	
	(0.717)	(0.717)	(0.717)	(0.261)	(0.261)	
Critic trial	0.114	−0.308	−0.308	−0.308		
	(0.631)	(0.466)	(0.466)	(0.466)		
CRT:Condition 3	−0.081	−0.076	−0.076			
	(0.977)	(0.975)	(0.975)			
CRT:Condition 3	−0.605	−0.603	−0.603			
	(1.215)	(1.212)	(1.212)			
Condition 2:Critic trial	−0.825					
	(0.828)					
Constant	−0.135	0.061	0.061	0.138	0.0003	−0.222
	(0.539)	(0.500)	(0.500)	(0.410)	(0.356)	(0.330)
Observations	486	486	486	486	486	486
Log Likelihood	−301.560	−302.056	−302.056	−302.190	−302.404	−304.127
Akaike Inf. Crit.	623.120	622.112	622.112	618.380	616.809	616.254
Bayesian Inf. Crit.	664.982	659.788	659.788	647.684	641.926	632.999
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table 7. Choice accuracy and CRT.

	Dependent variable:						
	Correct						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
time_norm	−1.350	−1.342	−1.449			2.446**	
	(1.822)	(1.813)	(1.819)			(1.115)	
fix_norm	−6.231**	−6.262**	−6.721**	−4.959**	0.254		
	(3.028)	(3.006)	(3.013)	(1.984)	(1.080)		
Condition 2	−0.299	−0.298					
	(0.266)	(0.266)					
Condition 3	−0.414	−0.413					
	(0.281)	(0.281)					
Critic trial	−0.268						
	(0.458)						
time_norm:fix_norm	36.149**	36.284**	37.118**				10.052**
	(15.076)	(14.961)	(15.018)				(4.713)
fix_norm:time_norm				27.874***			
				(9.130)			
Constant	1.066**	0.946**	0.778*	0.497*	0.527*	0.074	0.301
	(0.513)	(0.466)	(0.450)	(0.274)	(0.283)	(0.319)	(0.260)
Observations	485	485	485	485	485	485	485
Log Likelihood	−301.362	−301.531	−302.691	−303.002	−308.673	−306.035	−306.319
Akaike Inf. Crit.	620.725	619.062	617.381	616.003	625.347	620.070	620.638
Bayesian Inf. Crit.	658.382	652.535	642.486	636.924	642.083	636.806	637.374
Note:	*p<0.1; **p<0.05; ***p<0.01						

Table 8.Choice accuracy and eye movements.

Appendix C

Table 1:

	Dependent variable:			
	Average dwell time			
	(1)	(2)	(3)	(4)
All CRTs	0.280	0.329**	0.345***	
	(0.171)	(0.165)	(0.118)	
Weller without	0.060	0.023		0.245**
	(0.164)	(0.159)		(0.114)
pos2	−0.021			
	(0.060)			
pos3	0.064			
	(0.061)			
perc2	−0.080	−0.080	−0.055	−0.085
	(0.084)	(0.084)	(0.073)	(0.083)
perc3	−0.088	−0.088	−0.027	−0.110
	(0.083)	(0.083)	(0.073)	(0.082)
All CRTs:perc2	−0.065	−0.065	−0.002	
	(0.146)	(0.146)	(0.104)	
All CRTs:perc3	−0.290**	−0.290**	−0.138	
	(0.145)	(0.145)	(0.103)	
Weller without:perc2	0.086	0.086		0.042
	(0.140)	(0.140)		(0.101)
Weller without:perc3	0.208	0.208		0.013
	(0.139)	(0.139)		(0.100)
Constant	5.355***	5.366***	5.374***	5.391***
	(0.092)	(0.088)	(0.074)	(0.087)
Observations	478	478	478	478
Log Likelihood	−25.326	−22.430	−20.645	−22.511
Akaike Inf. Crit.	78.653	68.860	59.290	63.022
Bayesian Inf. Crit.	137.027	118.896	96.817	100.548
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 2:

	Dependent variable:			
	Number of fixations			
	(1)	(2)	(3)	(4)
All CRTs	0.585	0.504	0.382	
	(0.459)	(0.449)	(0.321)	
Weller without	−0.255	−0.150		0.195
	(0.440)	(0.433)		(0.316)
pos2	0.252			
	(0.183)			
pos3	0.249			
	(0.185)			
perc2	−0.277**	−0.277**	−0.181	−0.245*
	(0.135)	(0.135)	(0.129)	(0.134)
perc3	−0.338**	−0.339**	−0.207	−0.313**
	(0.135)	(0.135)	(0.129)	(0.134)
All CRTs:perc2	0.275*	0.275*	0.549***	
	(0.157)	(0.157)	(0.112)	
All CRTs:perc3	0.224	0.224	0.592***	
	(0.156)	(0.156)	(0.112)	
Weller without:perc2	0.359**	0.359**		0.536***
	(0.144)	(0.144)		(0.103)
Weller without:perc3	0.484***	0.485***		0.630***
	(0.144)	(0.145)		(0.103)
Constant	2.585***	2.719***	2.684***	2.749***
	(0.249)	(0.239)	(0.202)	(0.241)
Observations	478	478	478	478
Log Likelihood	−2,155.421	−2,156.667	−2,162.722	−2,159.536
Akaike Inf. Crit.	4,336.842	4,335.333	4,341.445	4,335.073
Bayesian Inf. Crit.	4,391.047	4,381.199	4,374.801	4,368.429
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 3:

	Dependent variable:					
	Choice Accuracy					
	(1)	(2)	(3)	(4)	(5)	(6)
All CRTs	−0.309	1.068	1.324*		1.826***	
	(1.089)	(0.809)	(0.751)		(0.536)	
All CRTs	2.183**	0.914		1.972***		1.632***
	(1.044)	(0.746)		(0.739)		(0.511)
perc2	−0.018		−0.669	0.217		
	(0.883)		(0.782)	(0.865)		
perc3	0.499		−0.107	0.599		
	(0.870)		(0.771)	(0.857)		
All CRTs:critic	0.029	0.091	−0.646		−0.723	
	(1.202)	(1.137)	(0.645)		(0.616)	
Weller without:critic	−0.698	−0.836		−0.643		−0.768
	(1.043)	(0.985)		(0.557)		(0.542)
All CRTs:perc2	3.133**		1.498			
	(1.456)		(1.043)			
critic:perc3	1.349		0.059			
	(1.422)		(1.002)			
Weller without:perc2	−2.261			−0.204		
	(1.381)			(0.990)		
Weller without:perc3	−1.864			−0.973		
	(1.372)			(0.975)		
Constant	−0.552	−0.384	0.012	−0.576	−0.228	−0.305
	(0.624)	(0.372)	(0.550)	(0.618)	(0.328)	(0.376)
Observations	486	486	486	486	486	486
Log Likelihood	−299.742	−302.680	−302.105	−303.735	−303.438	−304.302
Akaike Inf. Crit.	625.485	619.360	622.211	625.471	616.875	618.604
Bayesian Inf. Crit.	679.906	648.664	659.887	663.147	637.806	639.535
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 4:

	Dependent variable:				
	Correct				
	(1)	(2)	(3)	(4)	(5)
All CRTs	1.328	1.260	1.737***	1.516	1.567***
	(1.692)	(0.791)	(0.537)	(1.695)	(0.534)
Average dwell time	2.790			2.401	2.503**
	(3.491)			(3.445)	(1.167)
Number of fixations	−2.214	−2.156	0.801		
	(3.832)	(3.779)	(1.041)		
pos2	−0.351	−0.369	−0.395	−0.364	−0.365
	(0.251)	(0.252)	(0.252)	(0.251)	(0.251)
pos3	−0.598**	−0.555**	−0.583**	−0.604**	−0.605**
	(0.266)	(0.264)	(0.264)	(0.265)	(0.262)
All CRTs:time_norm	−0.706			0.169	
	(5.417)			(5.357)	
All CRTs:fix_norm	4.303	4.868			
	(6.107)	(5.997)			
All CRTs:critic	−0.667	−0.730	−0.718	−0.654	−0.655
	(0.612)	(0.618)	(0.615)	(0.611)	(0.611)
Constant	−0.453	0.309	0.047	−0.545	−0.574
	(1.052)	(0.484)	(0.362)	(1.048)	(0.472)
Observations	478	478	478	478	478
Log Likelihood	−292.255	−294.457	−294.787	−292.572	−292.573
Akaike Inf. Crit.	606.509	606.915	605.574	603.144	601.145
Bayesian Inf. Crit.	652.375	644.441	638.931	640.671	634.502

Note: *p<0.1; **p<0.05; ***p<0.01

References

[1] R. H. Thaler, Commentary—Mental Accounting and Consumer Choice: Anatomy of a Failure, Marketing Science 27 (1) (2008) 12–14.

[2] G. Antonides, R. Ranyard, Mental accounting and economic behaviour, Economic psychology (2018) 123–138.

[3] A. Tversky, D. Kahneman, The framing of decisions and the psychology of choice, science 211 (4481) (1981) 453–458.

[4] A. Tversky, D. Kahneman, Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment., Psychological review 90 (4) (1983) 293.

[5] R. H. Thaler, Mental accounting matters, Journal of Behavioral decision making 12 (3) (1999) 183–206.

[6] P. W. Henderson, R. A. Peterson, Mental accounting and categorization, Organizational Behavior and Human Decision Processes 51 (1) (1992) 92–117.

[7] S. T. Fiske, M. A. Pavelchak, Category-based versus piecemeal-based affective responses: Developments in schema-triggered affect. (1986).

[8] A. E. Singer, M. S. Singer, G. Ritchie, Role of transactions in mental accounting, Psychological reports 59 (2) (1986) 835–838.

[9] P. Jha-Dang, A review of psychological research on consumer promotions and a new perspective based on mental accounting, Vision 10 (3) (2006) 35–43.

[10] C. Heath, J. B. Soll, Mental budgeting and consumer decisions, Journal of consumer research 23 (1) (1996) 40–52.

[11] H. M. Shefrin, R. H. Thaler, The behavioral life-cycle hypothesis, Economic inquiry 26 (4) (1988) 609–643.

[12] N. Bonini, R. Rumiati, Acceptance of a price discount: The role of the semantic relatedness between purchases and the comparative price format, Journal of Behavioral Decision Making 15 (3) (2002) 203–220.

[13] N. Bonini, K. Tentori, R. Rumiati, Contingent application of the cancellation editing operation: The role of semantic relatedness between

risky outcomes, Journal of Behavioral Decision Making 17 (2) (2004) 139–152.

[14] A. M. Collins, E. F. Loftus, A spreading-activation theory of semantic processing., Psychological review 82 (6) (1975) 407.

[15] D. E. Meyer, R. W. Schvaneveldt, Meaning, memory structure, and mental processes, Science 192 (4234) (1976) 27–33.

[16] W. A. Bousfield, The occurrence of clustering in the recall of randomly arranged associates, The Journal of General Psychology 49 (2) (1953) 229–240.

[17] E. Tulving, Subjective organization in free recall of" unrelated" words., Psychological review 69 (4) (1962) 344.

[18] H. Estelami, Strategic implications of a multi-dimensional pricing environment, Journal of Product & Brand Management 12 (5) (2003) 322–334.

[19] H. Estelami, The effect of price presentation tactics on consumer evaluation effort of multi-dimensional prices, Journal of Marketing Theory and Practice 11 (2) (2003) 1–16.

[20] H. M. Kim, L. Kachersky, Dimensions of price salience: a conceptual framework for perceptions of multi-dimensional prices, Journal of Product & Brand Management 15 (2) (2006) 139–147.

[21] R. Thaler, Mental accounting and consumer choice, Marketing science 4 (3) (1985) 199–214.

[22] M. D. Johnson, A. Herrmann, H. H. Bauer, The effects of price bundling on consumer evaluations of product offerings, International Journal of Research in Marketing 16 (2) (1999) 129–142.

[23] A. T. Abraham, R. W. Hamilton, When Does Partitioned Pricing Lead to More Favorable Consumer Preferences?: Meta-Analytic Evidence, Journal of Marketing Research 55 (5) (2018) 686–703.

[24] I. F. Quotidiano, Antitrust, fine of 678 million euro to car and financial companies: "A cartel to manage loans" (January 2019). URL https://www.ilfattoquotidiano.it/2019/01/09/antitrust-multa-da-670-milioni-di-euro-a-case- automobilisitiche-e-finanziarie-un-cartello-per-gestire-i-finanziamenti/4884197/

[25] B. Argyle, T. Nadauld, C. Palmer, Real effects of search frictions in consumer credit markets (2017).

[26] S. Kulaev, Nearly half of mortgage borrowers don't shop around when they buy a home (January 2015). URL https://www.consumerfinance.gov/about-us/blog/nearly-half-of mortgage-borrowers-dont-shop- around-when-they-buy-a-home/

[27] J. Oechssler, A. Roider, P. W. Schmitz, Cognitive abilities and behavioral biases, Journal of Economic Behavior & Organization 72 (1) (2009) 147–152.

[28] E. I. Hoppe, D. J. Kusterer, Behavioral biases and cognitive reflection, Economics Letters 110 (2) (2011) 97–100.

[29] J. M. Liberali, V. F. Reyna, S. Furlan, L. M. Stein, S. T. Pardo, Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment, Journal of behavioral decision making 25 (4) (2012) 361–381.

[30] K. E. Stanovich, R. F. West, Individual differences in reasoning: Implications for the rationality debate?, Behavioral and brain sciences 23 (5) (2000) 645–665.

[31] D. Kahneman, Maps of bounded rationality: A perspective on intuitive judgment and choice, Nobel prize lecture 8 (2002) 351–401.

[32] T. S. Tullis, Designing a menu-based interface to an operating system, in: ACM SIGCHI Bulletin, Vol. 16, 1985, pp. 79–84.

[33] K. Frederickson-Mele, Usability testing an intranet prototype shell, in: Web Usability Workshop at Chi, Vol. 97, 1997.

[34] T. S. Tullis, Using card-sorting techniques to organize your intranet, Intranet Journal of Strategy & Management 1 (2003) 1–9.

[35] T. Tullis, L. Wood, How many users are enough for a card-sorting study, in: Proceedings UPA, Vol. 2004, 2004.

[36] K. Deibel, R. Anderson, R. Anderson, Using edit distance to analyze card sorts, Expert Systems 22 (3) (2005) 129–138.

[37] S. Frederick, Cognitive reflection and decision making, Journal of Economic perspectives 19 (4) (2005) 25–42.

[38] M. E. Toplak, R. F. West, K. E. Stanovich, Assessing miserly information

processing: An expansion of the Cognitive Reflection Test, Thinking & Reasoning 20 (2) (2014) 147–168.

[39] C. Primi, K. Morsanyi, F. Chiesi, M. A. Donati, J. Hamilton, The development and testing of a new version of the cognitive reflection test applying item response theory, Journal of Behavioral Decision Making (2015).

[40] K. S. Thomson, D. M. Oppenheimer, Investigating an alternate form of the cognitive reflection test, Judgment and Decision Making 11 (1) (2016) 99.

[41] A. Newell, P. S. Rosenbloom, Mechanisms of skill acquisition and the law of practice, Cognitive skills and their acquisition 1 (1981) (1981) 1–55.

[42] Weller JA, Dieckmann NF, Tusler M, et al. Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. Journal of Behavioral Decision Making. 2013; 26(2):198–212.

[43] W. H. Boom, A. Garde, N. O. Akseli, The European Unfair Commercial Practices Directive: Impact, Enforcement Strategies and National Legal Systems, Routledge Taylor and Francis group, 2014.

STUDY 3

CHAPTER 4

Cognitive Reflection and gaze behaviour in visual tasks

ABSTRACT

The cognitive reflection test (CRT) measures the ability to suppress an intuitive answer that easily comes to mind. The relationship between the CRT and different cognitive biases has been widely studied. The aim of this paper is to ask if cognitive reflection is related to how participants allocate their attention in two different visual tasks that involve overriding a prepotent response. We found that higher cognitive reflection scores predicted fewer errors when eye movements need to move away from a salient target and also a larger tendency to switch attentional settings depending on which of two stimulus colours provided more information. Our results suggest that cognitive reflection is related not only to patterns in our thinking but to how we allocate attention to the environment.

Keywords: *Cognitive reflection; visual attention; eye movements; antisaccade; visual search.*

1. Introduction

Cognitive reflection is the tendency to reflect on a question instead of reporting the first, potentially erroneous, response that comes to mind. It was initially measured by the Cognitive Reflection Test the popular three item test in which people have to solve three problems that automatically generate intuitive responses ([1]). All the questions in the CRT evoke a response that is incorrect but immediate, intuitive, prepotent, or automatic. This ability "to resist reporting the response that first comes to mind”([1], pp.35) has been interpreted in the literature using dual-system theories (e.g. [2]; [3] [4]; [5]; [6]). Errors on the CRT are believed to occur because System 2 (slow, effortful and reflective) fails to monitor System 1's outputs and to override its functioning (quick, intuitive and heuristic). As such, performance on the CRT is generally taken to indicate the degree to which a given person relies on System 1 in their thinking ([7]).

Cognitive reflection has been shown to predict rational thinking, reasoning ability across a wide range of heuristics and biases ([7]; [8]), decision-making skills, time and risk preferences ([1]), and thinking dispositions ([9]) which in turn are related with economic behaviour and decision. For example, Noori ([10]) showed that people with lower cognitive reflection are significantly more likely to exhibit the conjunction fallacy, illusion of control, overconfidence, base rate fallacy, and conservatism.

The relationship between cognitive reflection scores and a susceptibility to a wide range of heuristics and biases is consistent with a common cause of over-reliance on System 1. While the connection between cognitive reflection and heuristic thinking has been quite well established, less is known about whether cognitive reflection is related to heuristic control of attention in perceptual tasks. The link between perception and cognition has been explored in great quantities of literature (e.g. [11], [12], [13], [14]). Although perceptual processing is often described as automatic ([15]), it can be highly contingent on cognitive control, as enabled by attention, which plays an important role in determining what information we gather from the environment ([16]). Two-systems views (stimulus-driven and goal-driven) have a long history in research on visual attention as well ([17], [18]), and to the extent that cognitive dispositions measured by the CRT reflect a general tendency towards heuristic control, they may not only affect how we think, but what we attend to when impulses and goals conflict. Whether or not cognitive reflection is predictive heuristic control of attention is important, as many real-world behaviours depend on the ability to exert flexible control over visual processing ([19]).

In the present paper, we tested the hypothesis that cognitive reflection can predict how participants allocate their visual attention in a visual search task ([20]) and in the anti-saccade task ([21]). We chose these tasks because in both there is a prepotent behavioural response: in the visual search task, a matching bias leads participants to look at the colour mentioned in the search question

([22]; [23]), even though this occasionally leads to more searching than necessary. In the anti-saccade task participants must overcome the reflex to look at a salient target ([21]); "participants must suppress the reflexive urge to look at a visual target that appears suddenly in the peripheral visual field and must instead look away from the target in the opposite direction [...] A crucial step involved in performing this task is the top-down inhibition of a reflexive, automatic saccade" ([[21]], pp. 218).

For the first time in the literature, these tasks were analyzed as a function of CRT scores. Our goal was to test the hypothesis that people with higher cognitive reflection have a better ability to override prepotent control over their attention when such control is required (as in the anti-saccade task) or simply beneficial (as in the search task) and, consequently, a better ability to switch behaviour depending on the best strategy to complete a task efficiently. Below, we describe these tasks in more detail.

1.1. Visual Search Task

Rajsic, Wilson, and Pratt ([20]) used a modified visual search task to examine whether visual attention was biased to "confirm" a target's identity in a task where other strategies were available. In this task, participants search for a target letter among visually similar letters (e.g., a p among b's, d's, and q's) and provide a yes/no answer to a question (e.g., "is the p red") asking whether or not the target letter is a particular colour.

Critically, because the target was always present, attending the letters in the "yes" colour or the "no" colour exclusively would provide all the information necessary for an answer. The task affords two colour-based search strategies (see Figure 1). The first is to simply attend letters with the colour mentioned in the question. The second is to search so as to minimize the number of letters searched (i.e., look at whatever colour is the most rare, and infer the colour by exclusion if the target is not in this set). Even though the colour asked about did not predict the colour of the target overall, participants showed a bias to attend the colour matching the question instead of adopting the second, flexible strategy. Although this inflexibility causes more searching than is strictly necessary, it presumably reflects the influence of a fast and simple search heuristic (matching bias [23]).

By varying how many letters in a given search display possess the colour in the search question, and (by exclusion) how many did not, Rajsic, Wilson, and Pratt were able to measure which colour – if any – participants were more likely to attend. They found that even though the question provided no information about the target's likely colour, attention was biased to that colour, even though this led to more letters being attended than was strictly necessary on some trials. Rajsic, Wilson, and Pratt interpreted this to mean that searchers were biased to seek confirmation of the question, providing a negative response when confirmation failed. Subsequent research has shown that this bias is caused by a tendency to rely on a visual template to guide attention ([23]) instead of guiding attention to the rarest colour (which provides the most information per letter). That is, participants tend to rely on what seems to be the cognitively simpler strategy, one where the colour that is attended is not contingent on the stimuli shown on a given trial.

Given the relationship between cognitive reflection and reliance on thinking heuristics ([1]; [24]; [8]; [10]; [25]), we expected that reliance on this attentional heuristic may be stronger in those who score lower on the CRT. Specifically, we predicted that differences in reliance on attentional heuristics would manifest specifically when the number of letters whose colour matched the question were in the majority. In this condition, reliance on a visual matching bias would lead attention to stimuli matching the colour in the question (e.g., to red letters in Figure 1, given the question "is the b red?"). On the other hand, more flexible attentional control would lead attention to stimuli whose colour mismatch the question, as these stimuli provide more information (only two letters must be identified at most before a response can be given; e.g., attending green letters in Figure 1 given the question "is the b red?"). With respect to the CRT, we hypothesize that:

- $HP_{1.1}$: People with lower cognitive reflection (below the median) will exhibit the matching bias over all colour set sizes, with their first fixations being most likely to go to a letter whose colour matches the rule (the one mentioned in the rule).

- $HP_{1.2}$: People with higher cognitive reflection will guide attention to the colour that provides the most information (i.e., lets them finish a search with fewer letters being identified). This means that when there are six letters with the matching colour and two letters with the mismatching colour, these participants will be less biased (i.e., more likely to attend "green" in Condition 6, Figure 1).

We analyzed the first fixation because it provides a measure of initial attentional allocation without needing to consider how participants integrate information during search ([20]).

1.2. Antisaccade

To assess the relationship between cognitive reflection and suppression of a prepotent visual response, we measured gaze behaviour in the anti-saccade task ([26]). In this task a participant is asked to make a saccade in the direction away from the stimulus (anti- saccade condition) or a saccade in the direction of the stimulus (pro-saccade condition). They have to fixate a cross in the middle of the screen and, after the instruction ("look away" or "look toward"), the stimulus is presented to one side of the target. In these tasks two mental processes are required: "the inhibition of triggering a reflexive saccade towards the stimulus, and the inversion of the visual vector, i.e., the amplitude of the stimulus from one hemifield to the other" ([[27]], pp. 429). Successful anti-saccade performance involves inhibiting the reflex to look at sudden target onsets. If differences in CRT reflect a broad reliance on heuristic processing, then lower CRT scores should be associated with more errors on anti-saccade trials, where visual heuristics conflict with goals. Such a relationship could reflect a common reliance on inhibition, as correct responses in the CRT may require one to stop the first intuition and correct responses on anti-saccade trials require the ability to inhibit an automatic response to look at the stimulus that appears. Specifically, we hypothesized that:

- HP_2: People with lower cognitive reflection will make more errors compared to the people with higher cognitive reflection, in the anti-saccade trials specifically. This is due to their inability to stop an automatic behaviour that drives them to look at a target that appears.

On the other hand, we expect that there is no difference among CRT scores on pro-saccade trials because in that specific trial there is not an automatic behaviour (such as to look at the stimulus that appears) to block.

1.3. Change detection task

As an additional variable of interest, we measured the visual working memory capacity of participants using a colour change detection task ([28]). Visual working memory (VWM) capacity has been found to be quite variable between individuals, and predicts aspects of attentional control ([29]; [30]; [31]). Furthermore, span measures of working memory capacity predict anti-saccade error rate ([32]) and selecting the more efficient sub-set in a visual search ([33]). As such, we anticipated that it could mediate any relationship between the CRT and participants' control over their visual behaviour in the search and anti-saccade tasks, as CRT scores have been linked to working memory capacity more generally ([7]).

2. Methods

2.1. Participants

54 university students from University of Trento (34 female, 20 male; mean age= 23.6 years, SD=4.5) participated to this study. We collected as many participants as our resources allowed to increase statistical power. The experiment was made incentive compatible; during the instructions we told participants that they will receive 3 euros for participation and between 3 and 11,5 euros on the basis of their performance in one of the tasks.

2.2. Materials

The experiment was run on a Dell computer with a 23 inch screen. Responses were collected with a standard USB keyboard. Eyetracking data was collected using an Eye Link 1000 Plus (SR Research) Binocular Tower Mount which provides a data acquisition at up to 2000 Hz, sampling from the right eye. Computer-based tasks were programmed in Matlab using the Psychophysics Toolbox ([34]; [35]; [36]). CRTs were collected using paper-and-pencil.

2.3. Stimuli and Procedure

Each participant, after reading the instructions, completed the tasks on the computer with eye tracking. In addition to the visual search task ([20]), the anti-saccade task ([26]), and the colour change detection task ([28]), they completed three other tasks as part of a larger research project. After this experiment section they completed the four different CRTs mentioned above, without any time limit, using paper-and-pencil.

2.3.1. Visual search task

This task consisted of 90 trials containing a random mixture of 15 trials in each of six conditions, fully crossing two factors (matching set size [2, 4, or 6 letters] and correct response [yes, no]). Each trial began with a 1 second blank screen, after which the search prompt was given. Search prompts contained that trial's target letter and a colour (e.g., "Is the p red?"), presented for 2 seconds. Target letters were randomly selected from the set [p, q, d, b], with the remaining three letters used as distractors, and colours were randomly selected from the set [red, green], such that either 2, 4, or 6 letters matched the colour in the prompt and the target letter either was (for "yes" trials) or was not (for "no" trials) the colour from the prompt. After the prompt, participants saw a 500ms screen with just a fixation mark in the center of the screen. After this, the search display appeared, with eight letters (seven distractors and one target letter, approximately 0.6◦ X 1.1◦) appearing evenly distributed on the circumference of a 10.5◦ radius imaginary circle centred on the fixation mark (Figure 1).

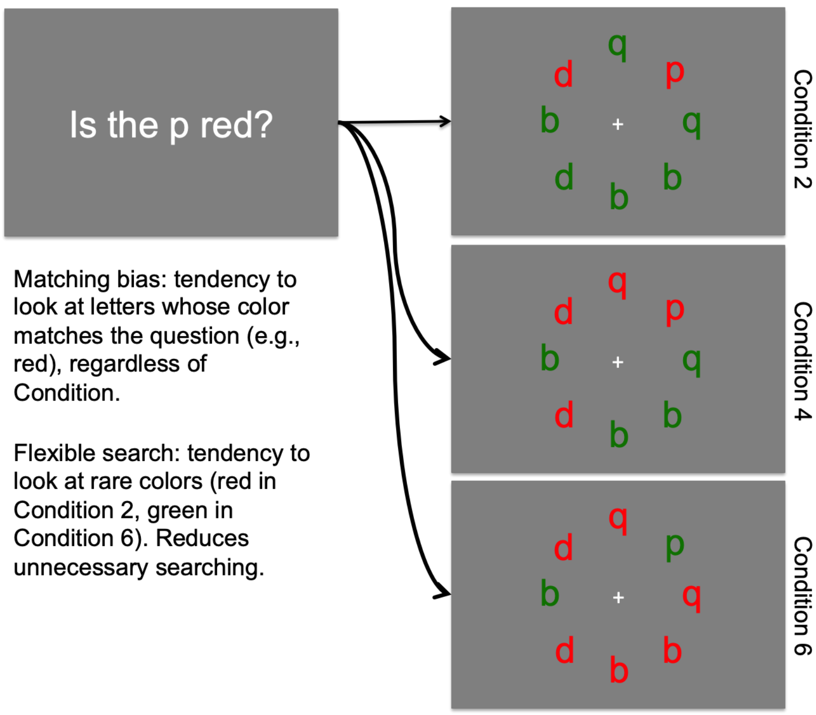

Figure 1. An illustration of the search task with sample search prompt and displays (not to scale). Note that the target letter (in this case, "p") is always present, but can be either colour (red/green) randomly.

This display remained on screen until the participant responded. Immediately following responses, the word "correct" or "incorrect" appeared in the center of the screen for 1s.

2.3.2. Antisaccade task

This task consisted of 80 trials; 40 pro-saccade trials and 40 anti-saccade trials, randomly intermixed. Each trials started with the instruction "Look at" or "Look away" printed in the center of the screen for 2 seconds to indicate what the appropriate response would be (Figure 2). This screen was replaced by a fixation cross and two small, white rectangles (0.3°), 9° to the left and right of fixation. After a randomized delay (evenly sampled from 2s - 2.5s), a 1.2° target rectangle at one of the two locations (randomly selected) flickered 3 times (50ms on, 50ms off) as shown in Figure 2.

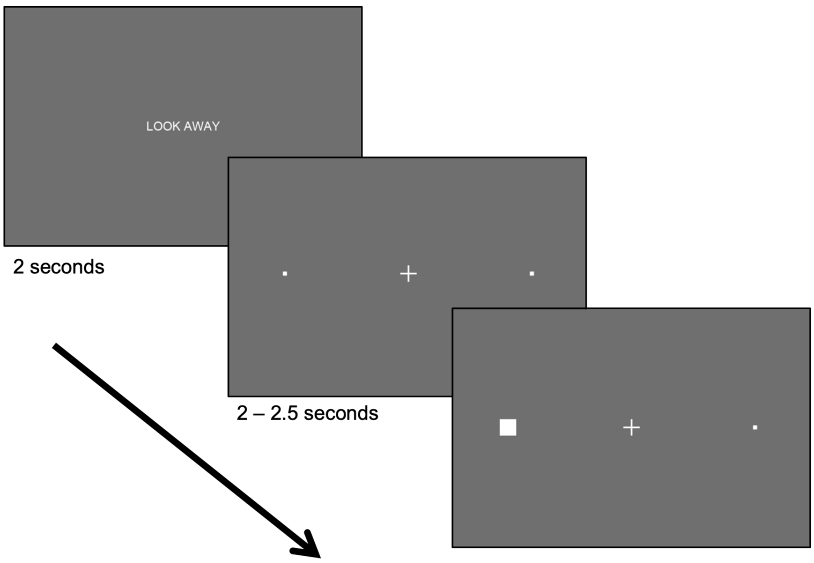

Figure 2. An illustration of an antisaccade trial. Note that the target flickered on onset.

When participants' gaze moved 3◦ left or right of fixation, this was considered a response. After 300ms, the next trial began unless an error was made. Errors were followed by 1 second of feedback ("You looked the wrong way!"). On trials where participants moved their eyes before the target, trials ended abruptly, and error feedback was presented for 2 seconds ("Do not move your eyes before the target occurs."). All such false-start trials were recycled so that every participant provided 80 trials with valid target responses.

2.3.3. Change detection task

This task consisted of 160 trials, fully crossing two factors (number of colours [2, or 6], and change [absent, present]. Each trial began with a one-second fixation mark on a blank screen. The memory sample display then appeared for 100ms, consisting of either 2 or 6 coloured squares, randomly placed in an imaginary square grid spanning 9.7◦, centred on fixation (Figure 3). Squares were 1.2◦ in height and width, separated by at least 2.4◦ (center to center). This display was replaced by a blank screen with a fixation mark for 900 ms. After this, the memory test display appeared until a response was entered. On change absent trials, the test display was identical to the sample display. On change trials, one square's colour was replaced with an unused colour. Colours were selected from a set of nine highly discriminable colours (red, green, blue, magenta, yellow, orange, grey, white, and black). Participants reported changes with the "A" key, and no change with the "L" key. A break was provided every 60 trials.

Memory capacity was calculated as k using the whole display equation ([37]) with hit rate and false alarm rate at set size six only. Data from set size two were used simply to verify that participants used the correct response mapping, and to reverse responses when false alarm rate exceeded hit rate.

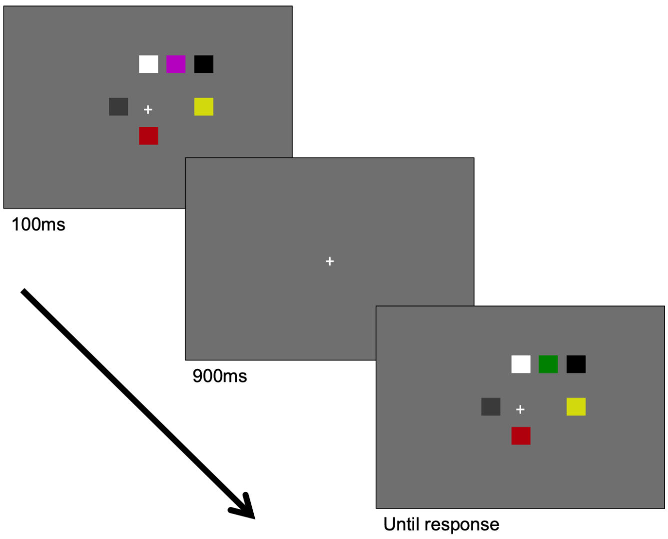

Figure 3. An illustration of the memory task, showing a "change" trial.

What we expected is that this working memory measure could predict whether or not participants exert more sophisticated control over their visual behaviour in the search and anti-saccade tasks. Given the relationship between CRT and working memory (r=0.33, [7]) we expect an interaction between these two factors in predicting visual attention in the visual search and anti-saccade tasks especially given that working memory ability has been shown to predict anti-saccade performance ([[38]]) and the ability to restrict visual search to a colour subset ([[39]]). It is important to note that our measure of working memory did not measure span, but rather capacity (but see [[40]]), and so generalizing these results should be approached with caution. Indeed, one might reasonably argue that this task better taps a visual short-term memory store than visual working memory per se, although researchers currently do not agree on how to best use these terms ([41]).

2.3.4. Cognitive Reflection Test

We used four different versions of the CRT in order to cope with the spread of the correct answers of Frederick's ([1]) and to have a more complete measurement of this construct. In addition to the original CRT ([1]), we used three tests ([8]; [42]; [43]).

Toplak et al. ([8]) added four items to the original test, to have a seven-item test (CRT7); Primi et al. ([42]) added three items to the original test (CRT long, CRT-L); Thomson and Oppenheimer ([43]) developed a four-item test in order to increase the number of questions in the test and to address a numeracy confounding.

The final CRT score was measured as the number of correct answers divided by the total number of questions.

3. Results

3.1. Task analysis

Before analyzing how the CRT predicts performance in our three cognitive tasks, we analyzed each individually to ensure that the expected patterns of results were found. Average performance in each task is shown in Figure 4.

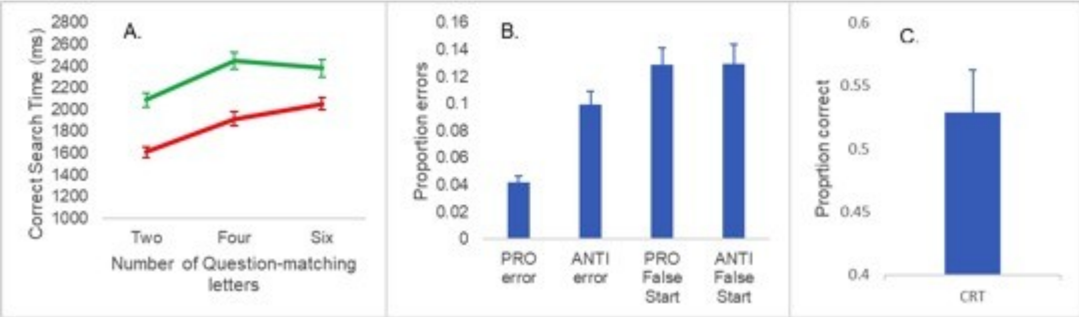

Figure 4. Average performance in the visual search (panel A), anti-saccade (panel B), and CRT(panel C) tasks. All error bars depict one standard error of the mean.

For the CRT, the average number of correct answers was 52.9% (SD = 24.4%). For the visual search task, as in previous studies the average correct search time was fastest for “yes” responses, $F(1, 53) = 126.70$, $p < .001$, $\eta^2_p = .71$. Search was also slower with more question-matching coloured letters, $F(2, 106) = 68.01$, $p < .001$, $\eta^2_p = .56$, with significant linear, $F(1, 53) = 95.60$, $p < .001$, $\eta^2_p = .64$, and quadratic, $F(1, 53) = 29.34$, $p < .001$, $\eta^2_p = .36$, trends. These factors also interacted, $F(2, 106) = 4.66$, $p = .011$, $\eta^2_p = .08$. The quadratic trend shows that participants did not exclusively attend to the “yes” coloured letters, but were nonetheless biased to attend them given that searches were faster when targets appeared in the small, matching subset than when they appeared in the small, mismatching subset, $t(53) = 12.57$, $p < .001$. At all set sizes, search time for “yes” responses was faster than search time for “no” responses, ts > 4.86, ps < .001. These results generally replicate previous findings that participants' attention is biased by the particular colour they are asked about, although this is indeed a bias and not rigid strategy. For the anti-saccade task, we found that participants made more errors on anti-saccade (M = 9.95%, SD = 7.2%) than pro-saccade (M = 4.2%, SD = 3.1%) trials, $t(53) = 6.09$, $p < .001$, confirming that anti-saccade trials led to impulsive responses. We also noticed that participants made a considerable number of saccades before the target appeared (which prematurely ended the trial). These false start trials were no more frequent in anti-saccade (M = 12.9%, SD = 10.9%) than pro-saccade (M = 12.9%, SD = 9.1%) trials, $t(53) = 0.07$, $p = .95$.

3.2. Hypothesis 1.1 and Hypothesis 1.2: Does CRT predict attentional flexibility?

We used generalized linear mixed models ([44]), which incorporated both fixed-effects parameters and random effects (subjects), to evaluate the hypotheses mentioned above. They allowed us to take advantage of trial-level data across all participants without collapsing all the data into sample averages. The best fitting model was defined, as supported by the model-selection method ([45]) as the one minimizing the Akaike information criterion AIC ([46]).

In order to test these hypotheses we used as a dependent variable a binary measure of whether the first fixation of each trial was on a matching colour (1) or if it was on a mismatching colour (0). We considered CRT score, matching set size, and change detection task scores as independent variables. Matching set size was entered as a dummy variable, with set size four as the intercept. This meant that the model was set up to predict increases or decreases in the probability of fixating a matching colour when the number of matching colours decreased from four (Condition 2), and when it increased from six (Condition 6).

As we can see from the Table 1 in the Appendix A, the interaction between CRT and matching set size six ("Condition 6" in the table) is significant in both models, the first (with only CRT: ($\beta_{INT\ 1} = -0.848$, $z = -2.613$, $p = 0.009$)), and the third model (with CRT and VWM capacity ("mem_cap" in the table): ($\beta_{INT\ 2} = 0.774$, $z = 2.307$, $p = 0.021$)) The beta coefficients of these interactions were negative, meaning that people with higher cognitive reflection looked first at the matching colour less often when matching coloured letters provided less information, compared to people with lower CRT. This provides evidence that greater cognitive reflection predicts better attentional flexibility: those with more cognitive reflection could attend the colour that reduced search load from their first eye movement (see Figure 5, rightmost panel). The interaction between matching set size two ("Condition 2" in the table) and the CRT was not significant, in either model (Figure 5, leftmost panel). This means that the CRT did not predict more frequent fixations on matching coloured letters when these letters were more salient. This could be due to the fact that these letters are salient for two reasons: they match the colour mentioned in the question being answered, and they are more informative. As such, both inflexible search, driven by matching bias, and flexible search, driven by information maximization, would drive attention to these stimuli, and so it is possible that high and low cognitive reflection participants attended these stimuli for different reasons. As we can see from the Figure 5, people with lower cognitive reflection look at the matching colour with an above-chance (purple line) probability, regardless of the number of letters with the matching colour (so regardless of the condition). People with higher cognitive reflection, instead, are able to switch strategy depending on the number of letters with the matching colour. Clearly, when attending the matching colour is useful (condition 2) or harmless (condition 4), higher CRT scores increase the probability of fixating a matching colour.

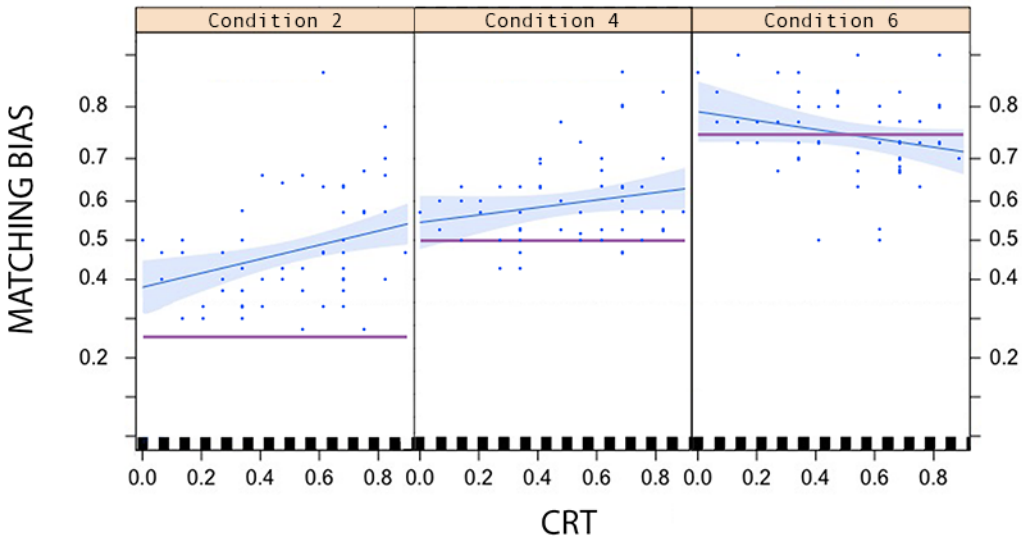

Figure 5. The probability of fixating a matching-colour letter first depending on the CRT in the three conditions. The horizontal purple line in each plot depicts the probability of fixating a matching-colour letter by chance alone, given the number of matching-coloured letters relative to the total number of letters on screen (eight) in each condition. Some participants had the same value; the value is adjusted to make overlapping data points visible.

However, when attending the matching colour is less helpful (condition 6), higher CRT scores are associated with a lower probability of fixating a matching coloured letter.

So, higher CRT scores are associated with more strategic control over eye movements in all the conditions which results in a matching bias in the second condition but not in the sixth where they looked more at the letters with mismatching colour).

We also computed a matching bias index ([22]), in order to measure how much a participant was biased to look at the matching colour (controlling for the chance). It is described by the following equation:

$$bias(p(match), chance) = \begin{cases} \frac{p(match)-chance}{1-chance}, & p(match) \geq chance \\ \frac{p(match)-chance}{chance}, & p(match) < chance \end{cases}$$

When the measured probability of inspecting the template-matching colour is greater than or equal to chance (.25, .5, and .75 for the matching subset sizes 2, 4, and 6, respectively), the extent of the matching bias is given by p(match) minus chance divided by 1 - chance, which expresses the degree to which fixations go to matching-colour letters above what would be expected by random eye movements. When the measured probability of inspecting the template-matching colour is lower than the chance, the matching bias is given by p(match) minus chance divided by chance.

As we can see from the Figure 6, while participants with lower cognitive reflection have a fairly consistent bias towards the matching colour, participants with higher cognitive reflection are better able to resist the matching bias and to use an attentional strategy that promotes switching which colour is attended from trial to trial in order to minimize the number of stimuli processed.

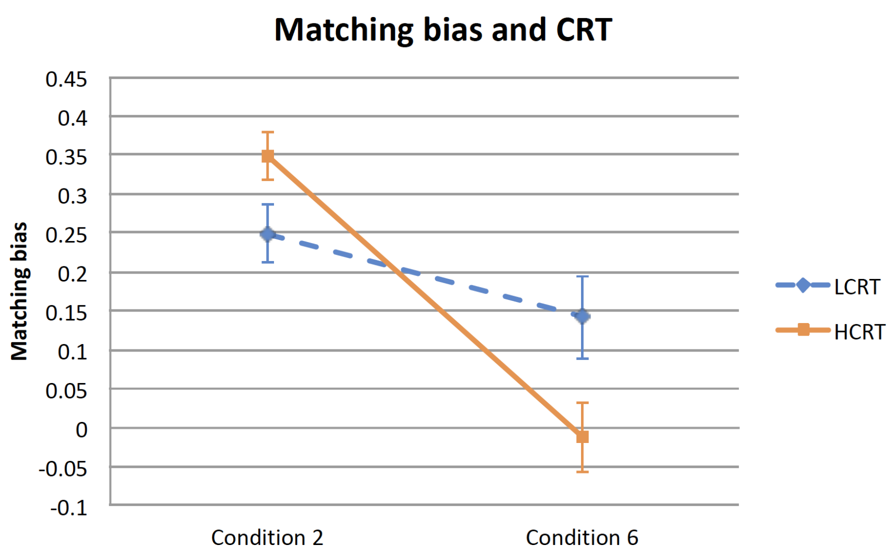

Figure 6. Matching bias index and CRT.

With regard to VWM capacity as a predictor as we can see from the Table 1 (Appendix A), it did not predict attentional flexibility in either model (the second with only the memory measure and the third with the memory measure and CRT). However, we note that the simple correlation between memory capacity and the matching bias index was positive in the matching set size two condition, r (52) = .299, p = .028, consistent with the conclusion that visual working memory capacity relates to the general ability to selectively attend information using colour.

These results verify hypotheses 1.1 and 1.2, lending support to the conclusion that cognitive reflection also predicts the ability to more flexibly control attention based on the current situation. With respect to the role of VWM capacity in this visual search task, we found only weak evidence for a relationship between it and this particular form of attentional control.

3.3. Hypothesis 2.1 and Hypothesis 2.2: Does CRT predict anti and pro errors in the antisaccade task?

To measure anti-saccade errors, we coded each trial's response for each participant as a 1 for correct and a 0 for incorrect saccades, excluding trials where eyes moved before the target onset. We considered CRT score and VWM capacity as independent variables.

We can see from Table 2 (Appendix A) only CRT scores (in the model with CRT as a single predictor) predict how often participants made errors in anti-saccade trials, (β_{AE} = 0.951, z = 2.099, p = 0.036). The negative beta shows that participants with lower cognitive reflection were more likely to make anti-saccade errors confirming hypothesis 2.2 (see Figure 7).

We computed a similar model on pro-saccade trials ("look towards" trials) to test if the CRT predicts the ability to inhibit a response (such as looking at a stimulus) or if it instead simply predicts better performance in both kinds of trials. As we can see from Table 3 (Appendix A), neither CRT nor visual memory capacity predicted performance in the pro-saccade trials of the anti-saccade task. As we can see from the Figure 7, the modelled probability of error in the anti-saccade trials goes from almost

0.14 for the participants that scored CRT=0 to 0.07 for the participants with CRT=1. So the probability to make an error in the anti-saccade trials is doubled if we compare participants with CRT=0 to participants with CRT=1. On the other hand, there is an almost equal, and very low, probability to make an error in the pro-saccade trials among the different CRT scores.

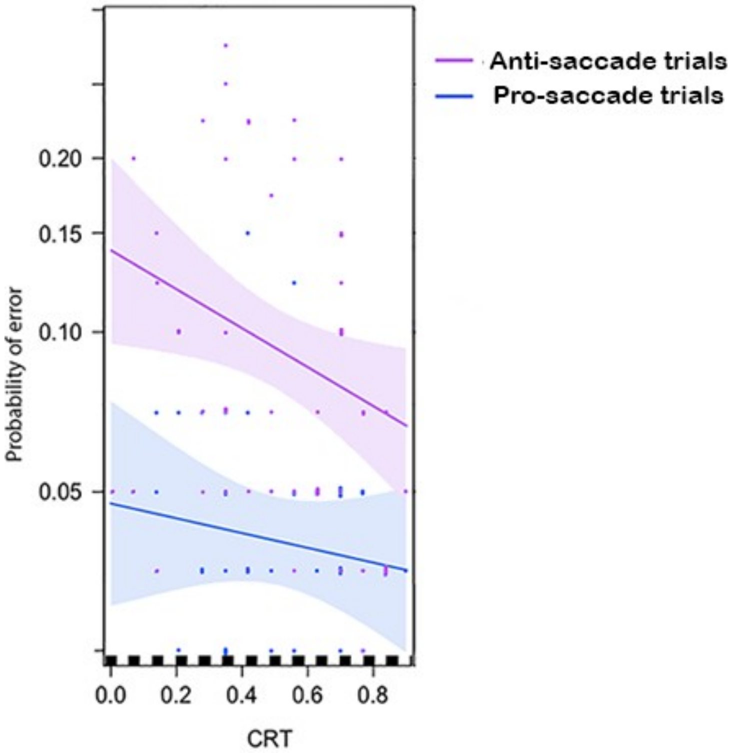

Figure 7. The probability of error depending on CRT and on the trial type. Several participants had the same value; the value is adjusted to make overlapping data points visible.

Indeed, if we compare the pro and anti-saccade trials (Figure 8) we can see that in the pro-saccade trials there is no difference between low CRT and high CRT participants; instead, there is a significant difference between these two groups in the anti-saccade trials (where it is more likely that people with lower CRT make errors).

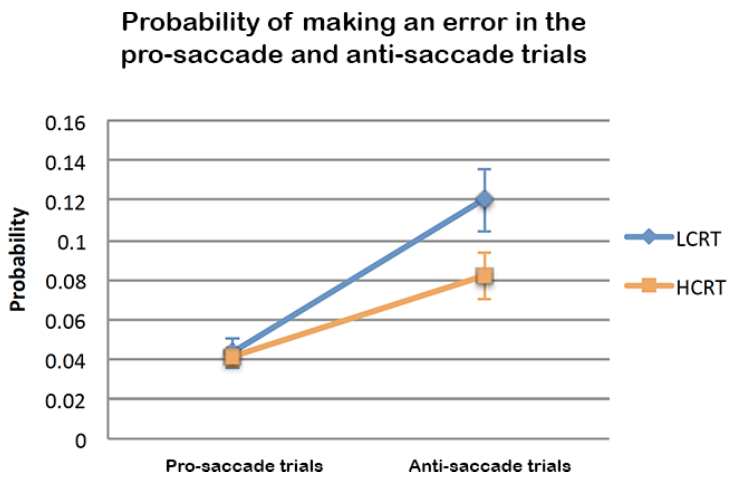

Figure 8. Probability of making an error in pro-saccade and anti-saccade trials depending on CRT.

As noted in section 4.1, we found a high false-start rate in the group with low cognitive reflection. We considered these saccadic false starts as a form of impulsivity so we decided to analyze if CRT predicts false starts in the anti-saccade task. To the extent that participants make impulsive eye movements before the target onsets, these too should be more frequent in participants with lower CRT scores. In order to test this hypothesis we used as dependent variable whether the participant made a false start (1) or not (0) on each trial. We considered CRT score and k score as predictor variables.

As can be seen in Table 4. (Appendix A), only CRT predicted false starts in the model with CRT as a single predictor,($\beta_{F\ S}$ = 1.247, z = 2.724, p = 0.006). The negative beta value shows that participants with lower cognitive reflection were more likely to make false starts. However, visual working memory capacity did not predict false starts in the anti-saccade task.

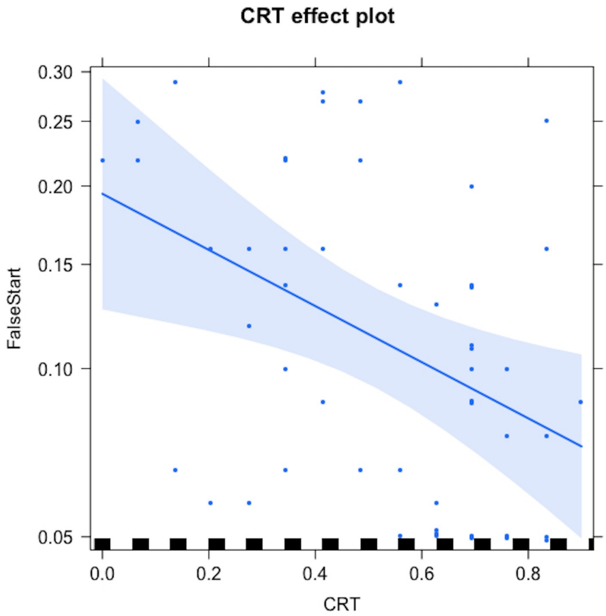

Figure 9. Probability of making a false start depending on the CRT.

As we can see from the Figure 9, the probability of making a false start in the anti-saccade task for participants with CRT=0 is equal to 0.2 and for participants with CRT=1 is around 0.07. Higher CRT scores, then, are associated with a higher rate of impulsive saccades in the anti-saccade task.

4. Discussion

Cognitive reflection is believed to result from individual differences in the control of System 2 processes over System 1 processes. The CRT was designed to exploit heuristics that lead to an erroneous but intuitive response, and scores on this task predict a wide range of heuristics and biases indicative of System 1 processes ([47]). In the present study, we sought to test whether cognitive reflection relates not only to reliance on heuristics in thinking, but also to reliance on heuristics in controlling attention in visual tasks. To do this, we measured CRT scores for participants who also completed an anti-saccade task (measuring the ability to resist attending to a salient input) and a visual search task (measuring the ability to flexibly change attentional settings on each trial). Both tasks pit heuristic actions (look at the target; attend the colour mentioned in the question) against more controlled responses (look away from the target; look at the colour that provides the most information).

We found that the CRT indeed predicted flexible attentional control in these two tasks. From these results, we suggest that cognitive reflection scores may not only indicate thinking dispositions, but potentially broader tendencies towards heightened behavioural control. Tentatively, we suggest that what links cognitive reflection and attentional control is the general tendency to suppress a planned or reflexive control signal and switch to another. Both tasks have an "intuitive" tendency that can be overcome with additional cognitive control that presumably relies on executive functions ([38]; [39]), which have been found to be stronger in those with higher CRT scores ([48]). In other words, CRT scores and attentional performance (as measured in the visual search and anti-saccade tasks) may both reflect differences in cognitive control. Although cognitive reflection may involve more than the suppression of the first answer ([49]), cognitive inhibition or impulsivity provides an intuitive link between high CRT scores and the improved attentional flexibility found in our experiment. Indeed, this executive function can be seen as a point of contact between cognitive reflection and attention differences; it could cause differences in attention and cognitive reflection. However, further research is needed to isolate the specific roles and causal relationships between these variables; while our results provide initial evidence that attentional control and cognitive reflection may be linked, further studies are needed in order to more clearly specify the precise overlap between the CRT and attentional control. It is worth being cautious in extending concepts like "inhibition" across tasks and domains (see [50] ; [51]). Rather, there may instead be specific executive functions that could overlap between attentional control and thinking that account for these relationships. Alternatively, these results could reflect a broader willingness or ability to adopt more cognitively demanding, but performance-enhancing, states of control.

Appendix A

Table 1: Matching Bias in the visual search task, as a function of CRT, visual working memory capacity and conditions.

	Dependent variable:		
	Matching Bias		
	(1)	(2)	(3)
CRT	0.387		0.199
	(0.246)		(0.515)
mem_cap		0.080	0.048
		(0.049)	(0.083)
Condition 6	1.148***	0.989***	1.280***
	(0.188)	(0.208)	(0.244)
Condition 2	−0.672***	−0.679***	−0.786***
	(0.173)	(0.194)	(0.226)
CRT:Condition 6	−0.848***		−0.774**
	(0.324)		(0.335)
CRT:Condition 2	0.343		0.283
	(0.300)		(0.310)
mem_cap:CRT			0.035
			(0.145)
mem_cap:Condition 6		−0.097	−0.058
		(0.065)	(0.067)
mem_cap:Condition 2		0.063	0.049
		(0.060)	(0.062)
Constant	0.185	0.150	0.083
	(0.141)	(0.156)	(0.266)
Observations	4,742	4,742	4,742
Log Likelihood	−3,039.823	−3,042.764	−3,036.876
Akaike Inf. Crit.	6,093.645	6,099.529	6,095.752
Bayesian Inf. Crit.	6,138.895	6,144.778	6,166.858
Note:	1	*p<0.1; **p<0.05; ***p<0.01	

Table 2: Errors in the "look away" trials in the antisaccade task, as a function of CRT and visual working memory capacity.

	Dependent variable:		
	Errors in the "look away" trials		
	(1)	(2)	(3)
CRT	−0.951**		−0.775
	(0.453)		(1.275)
mem_cap		−0.085	−0.023
		(0.091)	(0.195)
CRT:mem_cap			−0.040
			(0.386)
Constant	−1.845***	−2.092***	−1.803***
	(0.255)	(0.292)	(0.605)
Observations	2,160	2,160	2,160
Log Likelihood	−685.506	−687.198	−685.399
Akaike Inf. Crit.	1,377.012	1,380.396	1,380.799
Bayesian Inf. Crit.	1,394.045	1,397.430	1,409.188
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 3: Errors in the "look towards" trials in the antisaccade task, as a function of CRT and visual working memory capacity.

	Dependent variable:		
	Errors in the "look towards" trials		
	(1)	(2)	(3)
CRT	−0.403		−1.661
	(0.437)		(1.254)
mem_cap		−0.005	−0.155
		(0.090)	(0.187)
CRT:mem_cap			0.396
			(0.374)
Constant	−2.915***	−3.110***	−2.446***
	(0.246)	(0.288)	(0.568)
Observations	2,160	2,160	2,160
Log Likelihood	−376.833	−377.252	−376.257
Akaike Inf. Crit.	759.666	760.503	762.515
Bayesian Inf. Crit.	776.699	777.537	790.904
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 4: False starts in the antisaccade task, as a function of CRT and visual working memory capacity.

	Dependent variable:		
	False Starts		
	(1)	(2)	(3)
CRT	-1.247^{***}		-1.419
	(0.458)		(1.304)
mem_cap:CRT			0.033
			(0.389)
mem_cap		-0.008	0.041
		(0.098)	(0.197)
Constant	-1.421^{***}	-2.057^{***}	-1.508^{**}
	(0.262)	(0.316)	(0.622)
Observations	5,049	5,049	5,049
Log Likelihood	$-1,958.417$	$-1,961.904$	$-1,958.237$
Akaike Inf. Crit.	3,922.834	3,929.808	3,926.473
Bayesian Inf. Crit.	3,942.415	3,949.389	3,959.108
Note:			*p<0.1; **p<0.05; ***p<0.01

References

[1] Frederick S. Cognitive reflection and decision making. The Journal of Economic Perspectives. 2005;19(4):25–42.

[2] Sloman SA. The empirical case for two systems of reasoning. Psychological bulletin. 1996; 119(1):3.

[3] Epstein S. Integration of the cognitive and the psychodynamic unconscious. American psychologist. 1994;49(8):709.

[4] Stanovich KE, West RF. Individual differences in reasoning: Implications for the rationality debate? Behavioral and brain sciences. 2000;23(5):645–665.

[5] Kahneman D, Frederick S. Representativeness revisited: Attribute substitution in intuitive judgment. Heuristics and biases: The psychology of intuitive judgment. 2002;49:81.

[6] Evans JSB. How many dual-process theories do we need? One, two, or many?. 2009;.

[7] Toplak ME, West RF, Stanovich KE. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. Memory & cognition. 2011;39(7):1275.

[8] Toplak ME, West RF, Stanovich KE. Assessing miserly information processing: An expansion of the Cognitive Reflection Test. Thinking & Reasoning. 2014;20(2):147–168.

[9] Mata F, Miranda DM, Sallum I, et al. Do general intellectual functioning and socioeconomic status account for performance on the Children's Gambling Task? Frontiers in neuroscience. 2013;7:68.

[10] Noori M. Cognitive reflection as a predictor of susceptibility to behavioral anomalies. Judgment and Decision making. 2016;11(1):114.

[11] Wertheimer M. Gestalt theory.. 1938;.

[12] Arnheim R. Visual thinking. Univ of California Press; 1969.

[13] Rock I. The logic of perception. Cambridge, MA: Bradford . 1983.

[14] Kanizsa G. Seeing and thinking. Acta psychologica. 1985;59(1):23–33.

[15] Kahneman D. Nobel Prize Lecture.'. Maps of Bounded Rationality: A Perspective on Intuitive Judgment and Choice,'December. 2002;8:2002.

[16] Simons DJ, Rensink RA. Change blindness: Past, present, and future. Trends in cognitive sciences. 2005;9(1):16–20.

[17] Jonides J. Voluntary versus automatic control over the mind's eye's movement. In: Long J, Baddeley A, editors. Attention and Performance IX; Hillsdale, NJ. Erlbaum Associates; 1981.

[18] Shiffrin RM, Schneider W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychological Review. 1977;84(2):127–190.

[19] Kramer MR, Porfido CL, Mitroff SR. Evaluation of Strategies to Train Visual Search Performance in Professional Populations. Current opinion in psychology. 2019;.

[20] Rajsic J, Wilson DE, Pratt J. Confirmation bias in visual search. Journal of experimental psychology: human perception and performance. 2015;41(5):1353.

[21] Munoz DP, Everling S. Look away: the anti-saccade task and the voluntary control of eye movement. Nature Reviews Neuroscience. 2004;5(3):218.

[22] Rajsic J, Wilson DE, Pratt J. The price of information: Increased inspection costs reduce the confirmation bias in visual search. The Quarterly Journal of Experimental Psychology. 2017;Available from: https://doi.org/10.1080/17470218.2016.1278249.

[23] Rajsic J, Taylor JET, Pratt J. Out of sight, out of mind: Matching bias underlies confirmatory visual search. Attention, Perception, & Psychophysics. 2017;79(2):498–507.

[24] Toplak ME, West RF, Stanovich KE. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. Memory & Cognition. 2011;39(7):1275–1289.

[25] Boban S, Edward JS, Maggie G, et al. Stress and Risky Decision Making: Cognitive Reflection, Emotional Learning or Both. Journal of Behavioral Decision Making. 2016;.

[26] Hallett PE. Primary and secondary saccades to goals defined by instructions. Vision research. 1978;18(10):1279–1296.

[27] Kennard C, Leigh RJ. Using Eye Movements as an Experimental Probe of Brain Function: A Symposium in Honor of Jean Büttner-Ennever. Elsevier; 2008.

[28] Luck SJ, Vogel EK. The capacity of visual working memory for features and conjunctions. Nature. 1997;390(6657):279.

[29] Vogel EK, McCollough AW, Machizawa MG. Neural measures reveal individual differences in controlling access to working memory. Nature. 2005;438(7067):500.

[30] Fukuda K, Vogel EK. Human variation in overriding attentional capture. Journal of Neuroscience. 2009;29(27):8726–8733.

[31] Unsworth N, Fukuda K, Awh E, et al. Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. Cognitive psychology. 2014;71:1–26.

[32] Unsworth N, Schrock JC, Engle RW. Working memory capacity and the antisaccade task: individual differences in voluntary saccade control. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2004;30(6):1302.

[33] Sobel KV, Gerrie MP, Poole BJ, et al. Individual differences in working memory capacity and visual search: The roles of top-down and bottom-up processing. Psychonomic Bulletin & Review. 2007;14(5):840–845.

[34] Brainard DH, Vision S. The psychophysics toolbox. Spatial vision. 1997;10:433–436.

[35] Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial vision. 1997;10(4):437–442.

[36] Kleiner M, Brainard D, Pelli D, et al. What's new in Psychtoolbox-3. Perception. 2007.

[37] Rouder JN, Morey RD, Morey CC, et al. How to measure working memory capacity in the change detection paradigm. Psychonomic Bulletin and Review. 2011;18(2):324–330.

[38] Unsworth N, Schrock JC, Engle RW. Working memory capacity and the antisaccade task: individual differences in voluntary saccade control. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2004;30(6):1302.

[39] Sobel KV, Gerrie MP, Poole BJ, et al. Individual differences in working memory capacity and visual search: The roles of top-down and bottom-up processing. Psychonomic Bulletin & Review. 2007;14(5):840–845.

[40] Shipstead Z, Harrison TL, Engle RW. Working memory capacity and the scope and control of attention. Attention Perception and Psychophysics. 2015;77:1863–1880. Available from: 10.3758/s13414-015-0899-0.

[41] Cowan N. The many faces of working memory and short-term storage. Psychonomic bulletin & review. 2017;24(4):1158–1170.

[42] Primi C, Morsanyi K, Chiesi F, et al. The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). Journal of Behavioral Decision Making. 2016;29(5):453–469.

[43] Thomson KS, Oppenheimer DM. Investigating an alternate form of the cognitive reflection test. Judgment and Decision Making. 2016;11(1):99.

[44] Bates D, Mächler M, Bolker B, et al. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:14065823. 2014;.

[45] Myung IJ, Forster MR, Browne MW. Guest editors' introduction: special issue on model selection . 2000.

[46] Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. Sociological methods & research. 2004;33(2):261–304.

[47] Jackson SA, Kleitman S, Howie P, et al. Cognitive abilities, monitoring confidence, and control thresholds explain individual differences in heuristics and biases. Frontiers in psychology. 2016;7:1559.

[48] Del Missier FD, Mäntylä T, Bruin WBD. Decision-making competence, executive functioning, and general cognitive abilities. Journal of Behavioral Decision Making. 2012;25(4):331– 351.

[49] Szaszi B, Szollosi A, Palfi B, et al. The cognitive reflection test revisited: exploring the ways individuals solve the test. Thinking & Reasoning. 2017;23(3):207–234.

[50] Caswell AJ, Morgan MJ, Duka T. Inhibitory control contributes to "motor"-but not "cognitive"-impulsivity. Experimental psychology. 2013;.

[51] Caswell AJ, Bond R, Duka T, et al. Further evidence of the heterogeneous nature of impulsivity. Personality and individual differences. 2015;76:68–74.

CHAPTER 5

Conclusion

This thesis explored the concept of Cognitive Reflection and its influence on judgment and decision-making related to different kinds of tasks. The main idea of this thesis is related to the predictive power of cognitive reflection on heuristics and biases. Given its relationship with the monitoring of System 2 on System 1 we studied how cognitive reflection affects the consumer decision-making process and how people allocate their attention in visual search tasks. Four are the main empirical findings of this thesis and they are chapter specific:

- we demonstrated that numeracy is not the only relevant cognitive ability that plays a role in problem solving scenarios with high numerical components and we have proposed the concept of "overlapping vulnerabilities" in order to depict the nature of the consumer where different forces (cognitive abilities) contribute during the decision-making process for the final choice (Study 1);
- we showed that cognitive reflection affects how we analyze information in consumer problem solving scenarios; people with lower cognitive reflection are more affected by heuristics and biases (a more superficial analysis) also in tasks with high numerical components (Study 1);
- we showed that people categorize attributes of the same product (or service) and use integration/ segregation processes to analyze information and consequently make decisions (Study 2);
- we demonstrated that people with lower cognitive reflection are more affected by this mental accounting process and therefore they do not integrate all the information which is useful to make an accurate decision (Study 2);
- finally we demonstrated that cognitive reflection affects also how we allocate our attention in visual search, supporting the idea that this cognitive ability is a trait related not only to cognition but also to the initial step of perceptual visual search (Study 3).