The Roles of Convergent, Divergent Thinking, and Contextual Focus during Scientific Reasoning: Birth of the “Z” Model

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Jean-Baptiste Quillien

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Keisha Varma, Advisor
Wilma Koutstaal, Co-Advisor

September, 2019
Acknowledgements

I would like to thank my committee members for their support, feedback and kindness:

Dr. Varma Keisha
Dr. Koutstaal Wilma
Dr. McComas Jennifer
Dr. Varma Sashank

I would also like to thank my friends for the long conversations and sharing of ideas as well as

The fantastic research assistants from the Koutstaal’s lab, for their precious help -- coding and running participants during those past long and stimulating years.

Thank you
Dedication

J’aimerais dédié cette thèse a mon ami et mentor Mr. Milan Kovacovic, sans qui je ne serais pas là. Grace à qui j’ai su explorer. Merci pour le chemin à venir, mon ami.

Ne pas dédié cette même thèse a un autre ami, qui a vu naître ma passion pour la psychologie de l’enfant derrière ses yeux de chiot étonné, et qui hier me voyait finir mes études supérieures derrière un regard mature et appaisé: Fridu. Merci pour le chemin passé, mon ami.
Abstract

The aim of this paper is to bridge the process of scientific reasoning with the field of cognitive science, and more specifically, the cognitive mechanisms involved during reasoning. This intent of bridging scientific reasoning with cognitive mechanisms gave birth to a new model: the “Z” model of scientific reasoning. This model integrates the traditional scientific reasoning steps while depicting the cognitive mechanisms and mental flexibilities at use during reasoning. The goal of this experiment was to test the “Z” model and thus investigate the role of divergent and convergent thinking during scientific reasoning. In addition, the “Z” model highlights the importance of Contextual Focus during scientific reasoning. Contextual Focus is defined as the cognitive shift between modes of thoughts. Contextual Focus was tested to investigate its predictive power on our specific measure of scientific reasoning (Bouncing Ball Reasoning Task; BBRT) and a broader measure of scientific reasoning (Lawson Test of Scientific Reasoning; LTSR). In addition, the predictive power over scientific reasoning performances of Intellect and Openness, the personality traits of interest, was also tested. First, we hypothesized that participants experimentally primed to think divergently should perform better during the exploration of the problem space during a scientific reasoning task (Phase 1 of BBRT). As predicted, participants in the divergent thinking group generated on average more hypotheses than the participants in the convergent thinking and the control groups. Secondly, we hypothesized that participants experimentally primed to think convergently should perform better during the exploitation of the evaluative space during a scientific reasoning task (Phase 2 of BBRT). As predicted, participants in the convergent thinking group displayed on average fewer categorical errors than the participants in the divergent thinking or the control groups. In addition, Contextual Focus was found to be a significant predictor of the overall performance in exploring the problem space of our specific scientific reasoning problem. Intellect score over broader scientific reasoning (LTSR) performance, Contextual Focus and Intellect were found to be significant predictors of broader scientific reasoning (LTSR) performance. Those findings can also be interpreted with broader cognitive science lenses. Given that complex mental tasks such as problem solving and critical thinking also require divergent and convergent thinking, future research should test whether the priming used during our experimental protocol also leads to an advantage on more general reasoning tasks.
# Table of Contents

Acknowledgments........................................................................................................i

Dedication......................................................................................................................ii

Abstract.......................................................................................................................iii

Table of Contents.......................................................................................................iv

List of Tables...............................................................................................................v

List of Figures..............................................................................................................vi

Chapter 1: Introduction...............................................................................................1

Chapter 2: Literature Review.......................................................................................6

  Scientific Reasoning....................................................................................................6

  Cognitive Mechanisms During Reasoning.................................................................25

  Personality and Thinking...........................................................................................42

  Model Z: An Integrated Perspective to Scientific Reasoning...................................44

Chapter 3: Research Questions....................................................................................58

Chapter 4: Rationale for Operationalizations and Modifications..............................61

Chapter 5: Methods ....................................................................................................73

Chapter 6: Results.......................................................................................................79

Chapter 7: Discussion.................................................................................................108

References..................................................................................................................133

Appendix A................................................................................................................144
List of Tables

Table 1. Sub-skills and Items Measured by the Lawson Test of Scientific Reasoning......67

Table 2. Descriptive Statistics for All Variables.........................................................81

Table 3. Pearson Correlations among Phase 1 and Phase 2 Performance, and Individual Differences Measures.................................................................82

Table 4. Descriptive Statistics for Number of Hypotheses by Time and by Conditions...86

Table 5. Descriptive Statistics for Number of Categories by Time and by Conditions ....89

Table 6. Descriptive Statistics for Number of Switches by Time and by Conditions.........91

Table 7. Descriptive Statistics for Number of Categorical Errors by Time and by Conditions..................................................................................................................95

Table 8. Descriptive Statistics for Number of Awarded Points by Time and by Conditions.........................................................................................................................98

Table 9. Descriptive Statistics for Final Score by Conditions.....................................99

Table 10. Categorization of Participants’ Explanations..............................................100

Table 11. Percentage of Participants by Conditions coming to Correct vs. Incorrect Evaluations of their Explanations...............................................................101

Table 12. Regression Model Predicting Phase 1 Performance.................................104

Table 13. Regression Model Predicting Phase 2 Performance ...................................105

Table 14. Regression Model Predicting LTSR..............................................................107
List of Figures

Figure 1. The keypad from BigTrak robot…………………………………………………8

Figure 2. Frames defining the problem space and hypothesis space search …………10

Figure 3. The “Alpha” model …………………………………………………………………14

Figure 4. Reconciling the current state with the goal state ……………………………15

Figure 5. The iCASA thinking framework ………………………………………………26

Figure 6. Representational map of the human lateral cortex …………………………29

Figure 7. The cortical substrate of the perception-action cycle ………………………30

Figure 8. Schema of the exploitation-exploration in different domains …………..32

Figure 9. Problem space exploration and evaluative space exploitation in the “Alpha” model………………………………………………………………………………46

Figure 10. The “Z” model of scientific reasoning …………………………………………47

Figure 11. Problem space exploration in the “Z” model ……………………………….49

Figure 12. Evaluative space exploitation in the “Z” model ……………………………52

Figure 13. Need for CF conditional to convergent and divergent thinking capacity …70

Figure 14. Elaborative reproduction task ………………………………………………71

Figure 15. RFFT Parts 1 through 5 (different dots arrangement)……………………72

Figure 16. Research design by conditions and time …………………………………….78

Figure 17. Number of hypotheses generated during Phase 1 per Condition………………85
Figure 18. Number of hypotheses generated, 100-second intervals.........................87
Figure 19. Number of categories generated, 100-second intervals.........................89
Figure 20. Number of switches between categories generated, 100-second intervals.....92
Figure 21. Number of categorical errors exhibited during Phase 2 per condition...........94
Figure 22. Number of categorical errors exhibited, 100-second intervals...................96
Figure 23. Points awarded for generated explanations, 100-second intervals.............98
Figure 24. Problem Space Exploration in the “Z” Model of Scientific Reasoning ......113
Figure 25. Evaluative Space Exploitation in the “Z” model of Scientific Reasoning ....116
Figure 26. The “Z” model of scientific reasoning .............................................120
Figure 27. Hypothesis Space Search: Exploration & Exploitation .........................129
Figure 28. Experimental Space Search: Exploration & Exploitation......................130
Figure 29. “Z” Model 2.0.............................................................................132
Chapter 1: Introduction

Scientific reasoning can be defined from two different angles: the holistic perspective, defined by the way science works and organizes its knowledge; or, the atomistic perspective, defined by the reasoning we are engaged in when practicing science. Let us start with the holistic view, then zoom in into the apropos of this paper. Science is a concerted activity of knowledge seeking, where finding the truth is the ultimate goal. Philosophers such as Descartes, Kant, Bacon, Hume, al-Haytham, and, more recently, Popper and Kuhn have helped understand and build the rigorous methodology we undertake when compiling knowledge about the world in general and the way we conduct science in particular. Theories are the bedrock of science, and from those theories, hypotheses are made and tested, sometimes corroborated, sometimes refuted. Popper (2002) was the first one to emphasize the importance of falsification to distinguish "true" science from "pseudo-science." If a corroborated hypothesis agrees with a specific theory, then the theory is reinforced; if a corroborated hypothesis disagrees with a given theory, then it is the iconoclastic nature of science, the theory needs to be modified, maybe even "destroyed," to give birth to a new one. This abstract depiction of how scientific knowledge is organized is not the heart of my focus in this paper but describes its structure. The analogy with Mandelbrot's "Bonhomme" – where the whole shape is formed by smaller
replications of itself – holds true when considering the similarity between the cognitive mechanisms described by the holistic view of scientific reasoning, and those involved in the atomistic description of the scientific reasoning process: in both perspectives, generation and iterative refinement of ideas are present. The main difference is that they exist at different levels of abstraction. The atomistic view of science corresponds to what is described by Kuhn as "normal science"(1970). Kuhn argues that "normal science" takes place within a paradigm, where scientists exploit the given theories in place. Scientific discovery, on the other hand, occurs when the accumulation of data that does not fit existing theories is reorganized within a new paradigm. This process of scientific discovery is also called a paradigm shift. The apropos of this paper is this process of "normal science," as described by Kuhn. This process of scientific reasoning can be depicted as a succession of different forms of mental search (e.g. hypothesis versus experimental space search). While the different steps, facets of scientific reasoning, or again specific types of reasoning (e.g. proportional reasoning, control of variables) needed to reason scientifically have been identified, the scientific reasoning literature does not integrate some more basic cognitive processes well known in the creativity literature such as convergent and divergent thinking. For example, the cognitive dualism of exploration/exploitation, that is key to understanding the steps involved in seeking scientific truth, is not represented or explicitly taken into account by traditional views of scientific reasoning. The aim of this paper is to bridge the process of scientific reasoning with the field of cognitive science, and more specifically, the cognitive mechanisms involved during reasoning. This intent of bridging scientific reasoning with cognitive mechanisms gave birth to a new model: the “Z” model of
scientific reasoning. The “Z” model integrates the traditional scientific reasoning steps while depicting the cognitive mechanisms and mental flexibilities at use during reasoning.

The motivations to develop an integrated model was triggered by the fields of problem-solving, creativity and education. The “Z” model could be used by researchers to investigate scientific reasoning and, more generally, problem solving as a map that would help further illuminate cognitive mechanisms at play when reasoning. One could argue that valid scientific reasoning and successful problem solving are comparable in regards to the mental steps and cognitive processes. In addition to the investigational benefit, this mapping of scientific reasoning from a cognitive perspective should help organize and categorize what has already been found in related sub-fields with an interest in problem-solving and reasoning in general, such as in cognitive neuroscience or personality psychology. Consequently, such an integrated model will help to better plan future investigations and more generally support our understanding of complex thinking by proposing a cohesive model to bridge the different cognitive science sub-fields. Further, generating models of scientific reasoning has not only a research purpose, but it also aims to improve education. Valid scientific reasoning and, more generally, critical thinking are central to building a “stronger” educated society. Scientific reasoning and critical thinking are very similar; if the former implies a more specific type of knowledge, the latter uses the same abstract sequence of mental search using the same cognitive mechanisms and, both are driven by the need for mental exploration and exploitation.

In this paper, I will introduce two different models of scientific reasoning. The first, model “Alpha”, was developed with the intent to combine and present previous findings
about the various mental spaces explored during scientific reasoning. It details each recognized and necessary step from formulating a valid scientific question to evaluating the results of an experiment. The “Alpha” model will be presented between a classic paper highlighting that scientific reasoning is a mental search enterprise, and the different identified types of reasoning necessary to valid scientific thinking. Cognitive mechanisms will be the second broader point of the literature review. The second model, the “Z” model aims to depict scientific reasoning through the lens of cognitive mechanisms. Therefore, we will present those cognitive mechanisms, starting with their broader, more abstract constructs, to hone our way to their more specific form. The third part of the literature review consists in presenting the “Z” model. Because the “Z” model's aim is to integrate the scientific reasoning process and its cognitive mechanisms, a review of the cognitive literature bridging the “Alpha” model with reasoning will be used to present the “Z” model.

While the overarching goal of this dissertation is to test empirically the “Z” model. This testing entails the investigation of the role of specific cognitive mechanisms, or mental tools, when someone is engaged in scientific reasoning.

The underlying sub-research goal of this dissertation is to find out what role is played by the complementary cognitive processes convergent and divergent thinking during scientific reasoning. We target two specific steps, or forms of mental search, in the scientific reasoning process: At first the generation of variables and hypotheses related to a scientific problem (i.e. problem space exploration) and then, the conclusion drawn from constraining information, such as for example experimentation results (i.e. evaluative space exploitation). In addition to those mental tools associated with mental
exploration and exploitation, the type of flexibility linking the exploration of the problem space with the exploitation of the evaluative space (i.e. Contextual Focus: shift between mental exploration and exploitation) is also under scrutiny. Finally, the question of individual traits is also considered within the context of scientific reasoning, suggested by some literature linking the mental tools (i.e. divergent and convergent thinking) with some specific personality traits (i.e. openness and intellect).

Answers to our research questions shall also have some application in education, critical, creative thinking as well as the Scientific Reasoning field.
Chapter 2: Literature Review

Scientific Reasoning

The scientific reasoning and its method have been studied and developed by many philosophers and scientists. Sir Francis Bacon’s work and many others have inspired the modern iterative process of the scientific method (i.e. from Theory to Observation to Hypothesis to Experiment to Result, and back to Theory). More recently, Klahr and Dunbar (1988) pioneered the investigation of scientific reasoning as a multiple mental space search, and more particularly the hypothesis formation and experimental space. We will first, consider scientific reasoning as a cognitive search consisting in searching different mental spaces, as suggested by Klahr and Dunbar’s classic work. Secondly, we will present a model we created, the “Alpha” model of scientific reasoning that was mainly derived from 1—Klahr and Dunbar’s work on mental search (1988) and 2 – the modern iterative scientific method and Varma and colleagues’ work on the facets of scientific reasoning (2013). Finally, we will introduce different types of reasoning that have been associated with valid scientific reasoning.

Scientific reasoning as a mental space search enterprise. "The successful scientist, like the successful explorer, must master two related skills: knowing where to look and understanding what is seen" (Klahr & Dunbar, 1988). This elegant introductive
quote to experimental design and hypothesis formation set the tone to start this inspirational and, considered by many, classic paper. On the one hand, experimental design and the experimental space search are consistent with the procedure concerning an observation related to the problem to be solved. On the other hand, hypothesis formation is related to the creative and evaluative process of theories. Despite the fact that an interaction between these two important steps was recognized to play a critical role during a scientific endeavor (Mitroff, 1974), Klahr and Dunbar rightly argued that the interaction between experimental design and hypothesis formation during scientific reasoning was not well documented. They proposed two studies that would require participants to generate hypotheses and experiments, evaluate the results, and, consequently, re-evaluate their hypotheses. To enhance the ecological validity of their experiment, Klahr and Dunbar used a robot tank, “BigTrak.” Participants were taught how to use different commands to control the robot using a keypad (Figure 1). Participants mastered some basic syntax of the BigTrak language and were able to input and decipher lines such as:

**CLR 5 7 3 15 HOLD 50 FIRE 2 8 GO**

**CLR:** Clear the memory of any previous command

5: Move forward 5 feet

7: Rotate counterclockwise ( ), 42 degrees, (because one unit corresponds to an increment of six degrees; 7*6=42).

3: Move forward 3 feet

15: Rotate clockwise ( ), 90 degrees

**HOLD 50:** Pause for 50.1 seconds, so pause for 5 seconds

**FIRE 2:** Fire 2 times
In Study 1, after mastering some basic commands of the BigTrak robot, participants' scientific reasoning endeavor was investigated. They were asked to discover a specific rule: "How does the Repeat (RPT) function work?" Before entering any program, participants were asked to clearly state their hypothesis (e.g. "The N after the RPT function will make the robot do N times the preceding command"). A think-aloud procedure was used to gather data while participants were pressing each button on the keypad. The experiment stopped after forty-five minutes or when the participants were certain of knowing how the RPT function worked.

The scientific reasoning effort here resides in the fact that the RPT function did not follow the same syntactic logic of some other pre-learned functions. It was unclear to the participants if a number associated with the RPT function meant repeat X number of times the protocol, or +X number of times. In fact, the RPT function worked as follows: RPT 1
meant to repeat the last command; more generally, RPT X: Repeat the last X\textsuperscript{th} commands preceding the RPT function.

Determining the role of N (i.e. RPT 3, RPT N) was the essence of the scientific reasoning in this experiment. The ambiguity resided in the fact that the number N after the RPT function could have played different roles. Klahr and Dunbar classified the data into “frames” (Figure 2). "Frames," or organization of knowledge given a specific context, are another way to represent the hypothesis space (Minsky, 1974). In the “Alpha” Model, it is argued that if "frames" belong to the hypothesis space (arrangement of variable combinations), they also belong to a broader mental search: problem space search, which consists of the generation of variables of interest, before combining them, to solve the problem. The Hypothesis Space, or "frames," is represented by the potential interpretations of the role of N associated with the RPT function. In this experiment, the participants stated (think aloud) their hypothesis about the RPT N function, they experimented to test their hypothesis by entering a program on the keypad (e.g. CLR 5 7 RPT 2), and evaluated if the “BigTrak” robot tank course of action matched their predicted hypothesis.
Figure 2. Frames defining the problem space and hypothesis space search

We learned some very informative elements about the hypothesis space search while discovering the “BigTrak” RPT function. Klahr and Dunbar found that whereas some hypotheses were generated using prior knowledge (e.g. linguistic/semantic assumptions), other hypotheses relied mostly on the feedback received (i.e. observations of the “BigTrak” actions after entering the command). Two different groups of participants emerged: The theorists who favored revisiting the hypothesis space before finding the correct solution and the experimenters who reached the right solution by emphasizing their search in the experimental space. Because the participants were asked to formulate a hypothesis before each experiment, some valuable information about participants’ strategy was unveiled. The theorists’ strategy consisted of generating a theory belonging to a frame (e.g. N role: selector), and then subsequently exploring the experimental space generating experiments testing the value of the frame. The experimenters exhibited two distinct phases in time. At
first, they explicitly generated hypotheses by exploring the frame like the theorists, but in a second phase, they mostly focused on generating experiments. Earlier in the study, experimenters’ change of hypothesis was typically directed by the negative feedback received (some would change earlier, some would also move into different frames – e.g. from N counter to N selector). Later, experimenters would not explicitly state a hypothesis anymore but rather focus on searching the experimental space and engage in some trial and error to reach the correct conclusion about the RPT function. In other words, experimenters favored a direct experimentation of the function RPT rather than an abstract reflection about its functioning, by formulating a new hypothesis every time. The nature of search displayed by the theorists could be depicted as more abstract and internal than the one entertained by the experimenters.

This first experiment demonstrated the existence of two types of mental space search, suggested by some participants relying more on the hypothesis space search (theorists), while others were relying more on the experimental space search (experimenters). In light of the results of experiment 1, the researcher's next legitimate questions concerned the theorists and experimenters' strategies. The new research question was: if, as suggested by the theorist’s approach, it would be possible to find the correct rule by just visiting the hypothesis space; Therefore, would this lead to spending less time searching the experimental space? To collect their data, the protocol was modified. During study 2, participants would formulate not one, but multiple hypotheses before engaging in any experimentation. If the correct rule were found by some participants, this would confirm that the theorists in study 1 had elaborated the correct frame by succeeding in
visiting the hypothesis space. This time participants would engage in some experimentation only after generating several hypotheses. The results confirmed the idea that participants who had generated the correct rule by successfully searching the hypothesis space spent less time searching the experimental space to understand if they were right. Klahr and Dunbar’s work posited another interesting finding: when comparing study 1 and study 2, they found that participants spent less time searching the experimental space in study 2. By being encouraged to devote more time searching the hypothesis space, participants spent, on average, less time searching the experimental space before finding the correct solution.

Klahr and Dunbar's findings influenced the development of a new model of scientific reasoning: "The “Alpha” Model." The hypothesis space search, as conceptualized by Klahr and Dunbar, belongs to a broader space in the “Alpha” model: the problem space. In the same vein, the experimental space, in conjunction with the feedback, belongs to another broader space named in the “Alpha” model: the evaluative space. We acknowledge here that Klahr and Dunbar did not make the distinction between hypothesis and problem space search, nor, between experimental and evaluative space search. On the one hand, searching the hypothesis space search involves the combining of variables while simultaneously exploring the broader problem space by generating new potential variables. On the other hand, the experimental space is associated with the formation of experiments which aim to test the previously generated hypotheses. Again, Klahr and Dunbar do not explicitly dissociate the feedback from the experiment space search. However, someone could successfully produce a valid experiment and, at the same time, fail to evaluate it. We argue here that the experimental space search associated with the feedback given after the
realization of these experiments describe another broader mental space: the evaluative space.

While the hypothesis and experimental search were found to be important mental spaces to be searched during scientific reasoning, the “Alpha” model illustrates the step by step search of the different mental spaces we are engaged in during valid scientific reasoning. To facilitate the clarity of the apropos, in this section I will illustrate each step of the scientific reasoning process using a specific example in the "illustration boxes." In each illustration box, I will use a more specific instance of John Snow's discovery of the vehicle of Cholera in 1854. The illustration box could be seen as a concrete illustration to help the reader – after being exposed to Snow's scientific endeavor – to make connections between the steps of the “Alpha” model and the reader's experience of scientific reasoning.

The “Alpha” model of scientific reasoning. The “Alpha” model (Figure 3) is inspired by Khlar and Dunbar's findings (1988) and Varma and colleagues’ model (2013). As previously discussed, Khlar and Dunbar posited that two types of mental search, hypothesis and experimental space search, take place during scientific reasoning. The model developed by Varma and colleagues, based on a critical literature review of scientific reasoning, is specific and segments the scientific reasoning process into five facets (i.e. Generating Hypotheses, Hypothesis Testing, Reasoning from Evidence, Drawing Conclusions, Coordinating Theory and Evidence).
Please note that the research proposed in this dissertation focuses particularly on the first step, the generation of variables and hypotheses pertaining to the problem at stake (problem space) and the last two ones, the evaluation of some experiment(s) and the conclusion(s) drawn from it (evaluative space).

Regarding the question "what triggers scientific reasoning?", two of the most prominent philosophers of science, Karl Popper and Thomas Kuhn, have comparable answers. Scientific reasoning is a game of problem-solving according to Popper (2002); while for Kuhn, scientific reasoning is seen as a puzzle-solving endeavor (1970). Solving complex problems, puzzle solving, is one of the greatest human capacities. Two classes of problems are recognized: well-defined vs. ill-defined (Pretz, Naples & Sternberg, 2003). Well-defined problems have well-known operators and variables (e.g. Tower of Hanoi, Chess), like the rules and pieces of a game of chess. Scientific reasoning usually does not take place within a well-defined paradigm, but rather, within an ill-defined one. Ill-defined problems do not have clear boundaries; the operators and variables are multiple and
sometimes unknown by the solver. This leads to the first step of our scientific reasoning framework: exploring the problem space. The influential work of Newell and Simon (1972) about problem spaces describes two different problem states, the current and the goal states. The aim of the solver consists in reducing the "distance" between the two states (Figure 4).

![Figure 4. Reconciling the current state with the goal state](image)

In the realm of scientific reasoning, the current state could be a puzzling observation or a scientific question (Illustration Box 1) and the goal state would be a correct answer to the question or a valid explanation for the puzzling observation. The distance between the current state and the goal state – how “far apart” they are from one another – is critical, not only because it will help the thinker assess the correctness of her reasoning, but it will also help her refine her thinking along the different steps of the reasoning process. In other words, the "current-goal state" appreciation could be compared to some instrument allowing the thinker to assess the progress of her reasoning.
Exploring the problem space consists of defining and generating the operators and variables that will allow the current state to lead to the goal state. The purpose of this mental search is to identify the elements (operators and variables) that could be related to the problem at stake (Illustration Box 2).

Illustration Box 1: Finding a problem, a puzzling observation

John Snow was a British physician, born in 1813. He was interested in better understanding the cholera disease.

He was not satisfied with the hypothesis positing that cholera’s spread was facilitated by miasma. In fact, this champion explanation, of the time, failed to fully explain why within the same location some would be infected and others not. In 1854, a cholera outburst, in Soho (London) would give J. Snow the means to test an alternative hypothesis.

Illustration Box 2: Exploring Problem Space

At the time, “bad air” or miasma was the “suspected” operator responsible for the spread of cholera. Other operators or variables could have also been considered: Food, water, domestic or farm animals, genetics, etc...

John Snow suspected water to be the variable responsible for the spread.

The second step of our scientific reasoning framework is the hypothesis space search. Klahr & Dunbar (1988) posit that during scientific reasoning individuals explore, serially, two different spaces: first the hypothesis space, then, the experiment space (steps two & three). The overarching goal of the hypothesis space search is generative (divergent
Divergent thinking will be described in the forthcoming section about cognitive mechanisms but, in brief, is the cognitive mechanism which consists in generating more than one idea, solutions, or hypotheses related to a problem. Scientists generate competing hypotheses to explain/solve the problem. The hypothesis search phase shall be divided into two separate parts, an exploration phase (requiring a divergent thinking process), and an exploitation phase (convergent thinking). Convergent thinking will be described in the forthcoming section about cognitive mechanisms but, in brief, is the cognitive mechanism which consists in honing an idea or group of ideas into the best solutions. These two parts of the hypothesis space search are respectively named: abductive reasoning and retroduction (Fann, 1970; Lawson, 2010). Abductive reasoning consists of generating hypotheses using the operators and variables identified during the problem space search, while retroductive reasoning is the mental testing/monitoring of those hypotheses with our current knowledge (Lawson, 2010) (Illustration Box 3). Retroduction could be compared to a mental experimentation/validation of the generated hypothesis.

**Illustration Box 3: Hypothesis Space Search**

**Abductive reasoning:** a scientist could have hypothesized that contact with domestic animals was the reason why cholera would spread.  
**Retroduction:** the same scientist would have mentally discarded this hypothesis because some people who died from cholera did not have contact with domestic animals!

John Snow hypothesized that **drinking water** was the vehicle for cholera spread. Water was the operator identified; drinking water is the association giving birth to the complete hypothesis (e.g. not just contact with the contaminated water, but the ingestion of it).
After this process of abductive reasoning – retroduction (i.e. hypothesis space search), only "valid" hypotheses are brought to the next phase: the experiment space search. The experiment space search is step three, after determining one or two competing hypotheses, the scientist will design an experiment to test these hypotheses. There are multiple ways to search the experiment space, this search mainly consists of a formal gathering of information about the current state and the goal state, testing the chosen hypothesis. The search of the experimental space, like the hypothesis space search, is again dependent upon an explorative and exploitative part. Many tools are at the scientist's disposal: quantitative study, qualitative study, controlling for variables, gathering information through testing, interviews etc. The exploration part of the experimental search consists in recruiting knowledge about What can I do? "What are the ‘tools’ that I have at my disposal (e.g. knowledge about experimental design) to test the hypothesis or hypotheses I chose to test?" After generating many ideas about "What can I do? How can I test the hypothesis (or hypotheses)? What are the tools I know to test my chosen hypothesis?”, the scientist enters an exploitation phase of the experimental space search. During the exploitation of the experimental space search, the scientist will take some constraints, often dictated by the environment, into account. Some of the methods she had in mind may not be adequate for a given context (e.g. too small of a sample size, no existing/reliable tool to measure the variable at stake). Consequently, the scientist will have to hone her experimental design, from the ways she had in mind during the exploration phase to the "viable" options identified after the exploitative phase. As always, the back
and forth from exploitation to exploration takes place, the honing process leads to the chosen experimental design. In summary, experiment space search consists in designing the most appropriate experiment to test the hypothesis previously generated (Illustration Box 4.)

**Illustration Box 4: Experimental Space Search**

What can I do? (tools): Identifying the victims of cholera and collecting information about what water they drank prior to be contaminated.

How should I do it? (constraints): Building a map based on the number of victims to help investigate if a water source (e.g. Broad Street pump) was at the epicenter of the cholera spread.

John Snow tested his hypothesis that **drinking water** was the vehicle for cholera spread by interviewing people in Soho about their drinking habit, and he also built a map to quantify the number and locations of deaths due to cholera.

The last and fourth step of the scientific reasoning framework is the exploitation of the evaluative phase, consisting in computing and interpreting the data collected from the experiment space search and, consequently, making the right conclusion about the results (Chinn & Brewer, 2001). If the feedback is positive (i.e. the “current” and “goal state” agree), then our hypothesis is corroborated; if the feedback is negative, then our hypothesis is rejected (Illustration Box 5).
Concerning the evaluative space exploitation and, more specifically the solution, we need to keep in mind that the absence of evidence is not synonymous with the absence of relationship (e.g. "I may not see the moon by my window tonight; however, that does not disprove its existence"). Because the notion of valid scientific reasoning is critical when confronted with negative feedback, the scientist must consider different causes of failure. First, the absence of positive results can be due to an error made during the experiment space search (e.g. error in measurement, "it was a cloudy night, the moon was hidden" or incompatibility between the construct to be measured and the operational definition used to measure it: "My definition of the moon was incorrect, I thought it was more like a kind of star, therefore I did not see it while it was in front of me!"); Or, it can be due to an error made during the evaluation phase (e.g. misinterpretation of the results, wrong statistical analysis, "I was not looking in the correct direction, I could not have seen the moon").

Illustration Box 5: Evaluative Space Exploitation

The Map: The map that was built by Snow revealed that most victims lived next to the Broad Street pump. This source of data is quantitative.

Interviews: After interviewing people in the neighborhood of the Broad Street pump; Snow finds out that workers from the brewery did not get infected (drinking beer); one of the cholera deaths that was outside the neighborhood actually drank the water from the Broad Street pump. Finally, while all the victims had drunk the water from the pump, no cases of cholera were observed in a workhouse nearby the pump. J. Snow found out that the workhouse had its own well. This source of data is qualitative.

The convergence of the data played in favor of J. Snow’s hypothesis; In September 1854, the authorities closed the Broad Street pump, resulting in the stop of the cholera spread.
type of mistake occurs most often during the exploitation of the evaluative space. Secondly, if the negative feedback is not due to step 3 (experiment space search) or step 4 (evaluation), it means that the issue comes from the problem space exploration (e.g. some important operators are missing in the hypothesis), or the hypothesis space search (e.g. the operators and variables considered are correct but the relationship between them is incomplete or just wrong, "because of the alignment of the earth and the moon phase, the moon could not be seen that night"). This type of mistakes is found during the exploration of the problem space.

In addition to the description of the framework, we argue that the order in which each step is performed follows a chronological sequence, except for the problem space exploration which happens concurrently with every other step of the scientific reasoning process. The problem space exploration is, consciously or unconsciously, always on the "back burner" of our mind. We constantly monitor the gap between the current state and the goal state (Chrysikou, 2006); and, more operators or variables can be thought of and incorporated at any time. Thus, our hypotheses and the following steps will have to be changed or adjusted adequately.

To conclude this part, we posit that valid scientific reasoning can be decomposed as a succession of searches in multiple mental spaces. While one step informs the succeeding one, the generation of variables and operators (exploration of problem space) is concurrent with the successive search composing the scientific reasoning process. In the case of negative feedback, this model is very useful to 1- describe the scientific reasoning process; and, 2- identify when and what type of mistakes occurred. However, this model
does not capture what happens between or even within those steps, in terms of cognitive mechanisms. The field of cognitive science offers some additional information about the cognitive mechanisms and, more specifically, the different modes of thought at play during reasoning. The new, more integrated, “Z” model aims to depict the scientific reasoning process more accurately by incorporating cognitive mechanisms related to reasoning. The “Z” model will be presented in the last section of this literature review; but before, describing the cognitive mechanisms that are integrated into this new model, let us review the identified types of reasoning that contribute to valid scientific reasoning.

**Different types of reasoning during Scientific Reasoning.** After inquiring about the different mental spaces concerned by scientific reasoning, we are going to acknowledge, in this part, the different recognized types of reasoning necessary to valid scientific reasoning. With the intent of facilitating cohesiveness, we will describe those types of reasoning within the scientific reasoning mental spaces they could be associated with. While scientific reasoning is sometimes reduced to a hypothetico-deductive enterprise (Lewis, 1988; Lawson, 2000), it is in fact, a little richer than a simple two-step process. Different types of reasoning have been identified and related to scientific reasoning. Even if the purpose of this research is to investigate the role of more basic and specific cognitive processes (i.e. convergent and divergent thinking) during scientific reasoning, it is important to mention and briefly mention more abstract types of reasoning here. In fact, the Lawson Test of Scientific Reasoning (Lawson, 1978; Lawson, 2000) was designed to include sub-set measures of the different types of reasoning needed during scientific reasoning, such
as proportional reasoning, control of variables, probability reasoning, correlational reasoning, and hypothetical-deductive reasoning. We included the LTSR in this research as a broader more general measure of scientific reasoning.

The specific type of reasoning related with control of variables is associated with experimental design. When designing an experiment, control for variables is critical to increase chances of making the right conclusion about, for example any causal effect (Kuhn, 2007). College students exhibited difficulties in implementing control of variables, relying on one variable only. This kind of failure results in only partial consideration of the experimental space, which consequently places the thinker at risk when drawing conclusions, because of an incorrect or incomplete evaluative space. Still on the manipulation of variables, Tschirgi (1980) found that the outcome of the problem investigated (“bad” vs. “good”) influenced participants, from sixth graders to adults, in the choice of the variables to manipulate. In fact, it was only when the outcome was negative (e.g. “the cake was bad”) that the participants would generate logical disconfirming test, by only varying the hypothesized variable responsible for the outcome (e.g. type of sweetener: “honey vs. sugar”).

Probabilistic reasoning is the type of reasoning useful before making inferences. While probabilistic reasoning is universal and used by most of us (e.g. the chances that X occurs based on Y), it may be challenging and consequently lead to the generation of wrong inferences. Tversky and Kahneman were among the first to show that people do not rely on calculus properties when engaged in probabilistic reasoning, but rather rely on heuristics. For example, they discovered the conjunction
fallacy (1983), result of violation of probability calculus: a 31-year-old woman, liberal and outspoken would be wrongly categorized as a feminist bankteller instead of a bankteller only. Tversky and Kahneman also revealed the framing effect (1985), in which people’s probabilistic reasoning is influenced by the way the problem is presented. For example, people would judge differently the same food depending on the label: 90% lean vs. 10% fat. In the realm of scientific reasoning, failure to reason probabilistically could result in generating wrong hypotheses, because focusing on a variable of minor importance; or it could also influence the evaluative search of the scientific reasoning process, by misinterpreting the data and drawing wrong conclusions.

*Correlational reasoning* is also a very important aspect of scientific reasoning. Failure to engage in valid correlational reasoning results in making incorrect conclusions because of failure to search the evaluative space, such as interpreting negative correlation (Erlick, 1966), or making inappropriate causal claims (Shaklee & Tucker, 1980). In addition, flawed correlational reasoning (e.g. illusory correlation) can also be the result of wrong problem space and hypothesis space search; in fact, pre-conceived beliefs about variables can affect students’ interpretation of their correlation (Kuhn, Amsel, & O’Loughlin, 1988).

*Hypothetical deductive reasoning* is a type of thinking that is critical when generating hypotheses. Lawson and colleagues have identified three stages of hypothetical-deductive (H-D) reasoning development: from an inability to test hypotheses, to the ability to test hypotheses for observable causal agents, and finally to the capacity to
test hypotheses for unobservable causal agents (2000). Failures of H-D reasoning, or inadequate H-D reasoning, would result in problems when generating hypotheses and designing experiments, and consequently the incapacity to generate a suitable solution.

Cognitive Mechanisms During Reasoning

In this second broader section of the literature review, the selected cognitive literature of interest encompasses findings from psychology and neuroscience with empirical links to reasoning. The overall structure of this section consists of moving from more abstract and general claims about cognitive mechanisms to more specific and operationally defined ones. First, the iCASA thinking framework describes what is thinking and what important factors to consider when investigating reasoning. Then, the cognitive mechanisms integrated by the “Z” model will be unveiled, from the perception-action cycle to the need for exploration-exploitation and their particular mode of thoughts (i.e. convergent and divergent thinking) along with the corresponding cognitive flexibilities (i.e. spontaneous flexibility, reactive flexibilities, and Contextual Focus).

The iCASA thinking framework. The iCASA thinking framework provides a broad overview of the cognitive elements of thinking (Koutstaal, 2012). The iCASA thinking framework (Figure 5) is a useful guide when examining cognition and, more broadly, when investigating thinking (i.e. "what is thinking?"). I will summarize here the main points of interest of the iCASA thinking framework to help the reader recognize the connection between the thinking framework and the cognitive mechanisms used during
reasoning. The iCASA thinking framework is critical to understanding the “Z” model; and, especially, the cognitive flexibility involved during thinking. As we saw previously in the case of scientific reasoning, we navigate different mental spaces (e.g. problem space, evaluative space); but, while searching those spaces other factors such as concepts and emotions also play a role in our cognition and flexible thinking capacity.

When thinking about reasoning, the first idea that comes to mind is "idea" itself. Koutstaal described the idea as being composed or decomposed into four elements: Concepts, Emotions, Perception, and Motivation/Goals. Each idea, consisting of those four elements, could be described as a pencil stroke drawing of a landscape near a shore. Sometimes the pencil stroke goes above or below the horizon, such as the way ideas go in and out of consciousness. Koutstaal refers to peak awareness (mountains or larger hills) as when the idea spikes into our consciousness. We can imagine that this spike, resembling a
mountain in our drawn landscape, may be due to significant focus or attention to an idea spiking in our consciousness. Ideas are sometimes just in our peripheral awareness, close to the consciousness threshold. For example, some ideas can be present in our mind without requiring, or us soliciting, much attention to them at a certain moment in time. Those ideas close to the awareness threshold may be represented in our painting metaphor by the beach and smaller rocks along the seashore. Some ideas are below the consciousness threshold, deeper in our minds comparable to the bottom of the ocean.

The “Z” model was also inspired by another essential element described by the iCASA thinking framework; the different loci describing the space where the elements composing the ideas (Concepts, Emotions, Perception, Motivation/Goals) are evolving in. In fact, the concepts, the emotions, the perceptions, and the motivations are moving, independently from one another, along two loci described by two different axes: 1) Levels of Specificity (ordinate) and 2) Degrees of Control (abscise).

During our reasoning the needs to zoom in and zoom out are omnipresent. The pendulum of specificity goes from highly abstract to highly specific. Let us choose concepts to illustrate the levels of specificity. When we think about cats, for example, we could categorize them in many different ways: Animal would be superordinate to mammalian, very abstract element of the semantic group; and, "Tuffy," a very specific element of the semantic group, would be considered subordinate to Siamese. More semantic elements would be present along this abstract – specific axis.

The second locus, illustrating the degrees of control, describes the different levels of control we display while reasoning. When we reason, we demonstrate various levels of
control, from highly controlled to highly automatic. The spontaneous level could be considered to be in the middle of the two extremes. Let us take the element of perception to illustrate the degrees of control. When we compare a novice with an experienced driver, we may notice that the degrees of control in their perceptions may differ significantly. In fact, the novice driver may have to remind himself to look in the mirror before any maneuver; he may also force himself to listen to the noise produced by the engine to know when to change gear when driving a stick shift. The experienced driver will exhibit different degrees of control; as a result of experience, she will display a more automatic perception. She will not have to remind herself to look in the mirror or devote high focus to the engine noise to change gears when required.

Finally, the dynamic movements of the core element of cognition (Concepts, Emotions, Perceptions, Motivations/Goals) are not only taking place in the mind, but also in the brain (Activity, Connectivity/Structure) and the environments (Physical, Social, and Symbolic). During cognition, brain, mind and environment are interwoven, giving individual differences such as personality traits (e.g. Intellect/Openness) a role to play. The iCASA thinking framework describes ideas and their components (Concepts, Emotions, Perceptions, Motivations/Goals) as evolving in a space defined by degrees of control and locus of specificity and, thus, in different spheres of our reality (Mind, Brain, and Environments). The iCASA Thinking Framework does not only provide a complex and valuable guidance to approaching cognition, but it also embodies the need for an integrated depiction of the different types of cognition such as scientific reasoning. Similarly to the
iCASA thinking framework, a more specific aim of the “Z” model is to bridge and integrate findings from different fields of study: Scientific Reasoning and Cognitive Science.

*Perception-action cycle.* As we previously discussed, the iCASA thinking Framework describes cognition happening between environment, brain, and mind. The perception-action cycle describes what happens between the environment and the brain (Fuster, 2015). The perception-action cycle comprises a communication network in our brain that allows us to perceive information from our environment and consequently act upon it. During a goal-directed sequence, the posterior regions of our brain are mainly devoted to the processing of the stimuli perceived from our environment (e.g. occipital lobe), while the anterior regions (e.g. prefrontal cortex) are responsible for encoding the actions to be made upon our environment (Figure 6). From a neuroscientific perspective, the perception-action cycle is defined by the interaction between the posterior and anterior cortices (Figure 7).

![Figure 6. Representational map of the human lateral cortex](image)
At the heart of the perception-action cycle lies the “Z” model of scientific reasoning. The developed model attempts to represent what happens cognitively between perception and action during scientific reasoning. It describes what individuals cognitively do after perceiving the problem and before deciding on an action/solution. In the creativity literature, the notion of making and finding is defined by our capacity to impose some top-down action to create, ultimately perceive, in a bottom-up fashion, the change in our environment caused by our actions (Koutstaal & Binks, 2015). The voluntary process of making is to be contrasted with the discovering process of finding. De facto, bringing to attention another important and closely related duet: exploration and exploitation. The corresponding shift between perception and action, making and finding, exploration and exploitation, are key to understanding the cognitive dynamics and the essence of the Model "Z." Now, as suggested by the iCASA thinking framework, let us move from the brain structure to the mind by having a closer look at exploration and exploitation.
Mental exploration and exploitation. The duet of exploration and exploitation is not idiosyncratic to cognition or more specifically mental search but also visual and physical search. Balancing between exploring and exploiting has been well studied and documented in the animal foraging literature (Hills et al., 2015). It is clear that globally exploring its environment is critical for an organism to find food to harvest, and then exploiting locally is the natural continuation of the first. The cost-benefit ratio triggers the tradeoff between the two: as the local exploitation goes on (time) the cost/benefit ratio increases until a certain threshold is reached, then a more global exploration is more attractive and so on and so forth. The animal foraging literature has even proposed a mathematical formula to understand how much time an animal should optimally spend exploiting before switching back to exploring the environment (Charnov, 1976). In our complex world of perpetual change, being able to mentally shift appropriately between exploration and exploitation is considered the key component of adaptive behavior (Cohen, McClure, & Angela, 2007). Many different domains are concerned with the exploitation-exploration tradeoff, from patch foraging to visual focus or, again, problem solving (Figure 8). The exploitation-exploration tradeoff is related to the degrees of control as described by the iCASA thinking framework. We focus our attention on a local part of the problem or defocus to a more global one. The notion of degrees of control in the exploitation-exploration tradeoff seems to be dependent upon the same cognitive mechanisms across domains. In a study where participants were primed to visually search (hidden treasure game) in a clustered fashion (favoring exploitation) vs. a diffuse manner (favoring exploration), scientists found that when performing an anagram task, participants who were
primed to exploit spent more time with each letter set (Hills, Todd & Goldstone, 2010). The effect of the priming decreased with time as the participants tried to solve more anagrams, but the fact that a transfer was observed from visual priming to a cognitive domain suggests that the same central executive was responsible for the control or shift between exploration and exploitation during visual and mental search. The “Z” model aims to integrate the need for exploration and exploitation. In the next section, more specific constructs, related to exploration and exploitation, are considered.

![Schema of the exploitation-exploration in different domains](image)

*Figure 8.* Schema of the exploitation-exploration in different domains

**Convergent and divergent thinking.** As we move along in this section from more abstract to a more specific construct, integrated by the “Z” model, one shall wonder: “what
are the cognitive tools or mechanisms related to exploration and exploitation?" Convergent and divergent thinking are two cognitive mechanisms or types of cognitive searches that are often linked to the field of creativity. Those two types of cognitive search are used during any complex reasoning (DeYoung, Flanders & Peterson, 2008; Guilford, 1956; Guilford, 1959; Hommel, 2012). Also referred as modes of thinking, they could be broadly defined as follows: Divergent thinking consists of producing many possible solutions to a given problem, while convergent thinking involves determining the most optimal solution to a problem. On the one hand, exploration of new ideas – requiring generation, defocus, and bottom-up processing – is associated with divergent thinking. On the other hand, exploitation of ideas – requiring honing, focus, top-down processing – is related to convergent thinking.

**Divergent thinking: An explorative mental instrument.** Divergent thinking can be portrayed as an internal, bottom-up, and generative cognitive mode of thought that occurs when we mentally search for multiple solutions to a problem (Chrysikou, 2019; Hommel, 2012). Divergent thinking is not a new concept. More than fifty years ago Guilford described this mode of thinking (1956). Divergent thinking is often perceived as an index for creative potential (Runco & Acar, 2012). Divergent thinking is related to associational reasoning and is the mode of choice when we generate ideas (Hommel, 2012). Traditionally, the definition of divergent thinking acknowledges four different elements (Guilford, 1959):

- Elaboration is the level of detail of each idea;
- Fluency is the capacity to generate numerous ideas rapidly;
• Originality is the ability to find novel and unique ideas;

• Flexibility is the number of different domains used in each idea.

Elaboration is the level of detail expressed in, or the specificity of, each idea generated. Elaboration will play different roles during divergent thinking depending on whether the task requires a more abstract versus a more specific type of search. When someone is performing a divergent thinking task when he has to search in a more abstract domain (e.g. discussing philosophy), the elaboration of his ideas will enhance his overall performance by allowing him to make more abstract connections between broader domains.

Originality and fluency are related elements. When measuring divergent thinking a strong relationship exists between fluency and originality (Hocevar, 1979). The more ideas an individual can generate the more original they tend to become with time (Beaty & Silvia, 2012).

The flexibility component of divergent thinking is the capacity to use or access different domains when generating or searching for ideas. This type of flexibility is also referred to as spontaneous flexibility, in opposition to reactive flexibility (Eslinger & Grattan, 1993). Spontaneous flexibility occurs when there are no constraints on the individual to generate ideas. Spontaneous flexibility and reactive flexibility will be surveyed in more detail later in this section (i.e. cognitive flexibilities). We posit that spontaneous flexibility, fluency, and originality are related because generating ideas in multiple domains should result in generating deeper and consequently more original ideas overall.
The cognitive neuroscience literature, and more especially brain network activation studies posit that the default-mode network (DN) is associated with divergent thinking search (Chrysikou, 2019; Beaty, Benedek, Keaufman & Silvia, 2015; Beaty et al., 2016). The default mode network is associated with imagination and "reconstructive mental activities," such as imagining future events, feelings, and the thoughts of others. The default-mode network is activated when we engage in flexible mental exploration or simulation when we, for example, try to predict the consequences of action on our environment. Divergent thinking, the generation of ideas through associational thinking, is dependent upon the default-mode network. The brain regions showing activity when the default-mode network is activated, are the ventromedial prefrontal cortex (processing emotional and self-related information), posterior cingulate cortex (mental imagery and autobiographical re-experiencing), and the right and left posterior parietal cortex (switching attention and attentional control). The default-mode network is also described as being related to internally-driven mental activity, when we, for example, generate hypotheses about an observed puzzling phenomenon.

**Convergent thinking: An exploitative mental instrument.** Convergent thinking is related to analytical reasoning, and it is the mode of choice when we hone or select specific ideas (Hommel, 2012). Convergent thinking is described by a more focused, top-down, and honing mode of thought (Chrysikou, 2019); for example, when we search for the best solution to a problem (Hommel, 2012). Convergent thinking shall be seen as the behavioral expression of the need for exploitation. Convergent thinking is traditionally defined as the cognitive mode requiring an individual to find the most optimal solution to a problem.
Convergent thinking has been found to be correlated with intelligence, while divergent thinking was not (Chermahini & Hommel, 2010). Yet, intelligence tests are mainly convergent thinking tasks, consisting in finding the one, and only one, correct answer. Convergent thinking is associated with top-down processing – requiring focus – and also with fluid intelligence, which suggests an active role played by Executive Functions during convergent thinking mode.

The cognitive neuroscience literature and more specifically brain network activation studies posit that the executive central network (ECN) is associated with convergent thinking search (Chrysikou, 2019; Beaty, Benedek, Silvia & Schacter, 2016). The executive control network is responsible for the planning and monitoring of our goal progress (Spreng, Stevens, Chamberlain, Gilmore & Schacter, 2010) (e.g. current and goal state during scientific reasoning). Focusing, coordination, and effortful control are the primary cognitive activities associated with the activation of the executive control network (Seeley et al, 2007). The topography of the executive control network was realized using different types of tasks such as creative story generation (Howard-Jones et al., 2005), piano improvisation (Limb & Braun, 2008), insight problem solving (Kounios, Fleck, Green, Payne, Stevenson, Bowden & Jung-Beeman, 2008), generation of novel analogies (Geake & Hansen, 2005), and design in the visual arts (Ellamil, Dobson, Beeman & Christoff, 2012). Koutstaal and Binks posit that such findings highlight the importance of "deliberate attentional and evaluative process to our creative endeavors" (2015). When convergent thinking helps us hone our ideas and evaluate their usefulness, the brain regions associated with the executive control network, its brain activation blueprint, are the right and left
dorsolateral prefrontal cortex (keeping track of goals and short-term working memory), the anterior cingulate cortex (evaluation of progress and monitoring errors), and the right and left posterior parietal cortex (attentional control and attentional switching). The executive control network is also described as being related to externally-driven purposeful mental activity, when we, for example, interact with our environment (e.g. perception-action cycle).

We mentioned earlier that divergent and convergent thinking were used during complex reasoning, and consequently during creative thinking. A creative outcome is defined as something being useful and original. Both cognitive mechanisms will play a different role: Divergent thinking will help us generating original ideas, while convergent thinking will assist us in maintaining the required degree of usefulness. The need to be mentally flexible within and between each mode of thought is critical. Both networks, the ECN and DN are associated with creative cognition (Beaty, Benedek, Silvia & Schacter, 2016; Beaty et al. 2018).

In the context of scientific reasoning we should expect divergent thinking to facilitate the exploration of the problem space and convergent thinking to facilitate the exploitation of the evaluative space. The next section introduces the cognitive flexibility related to the shift between convergent and divergent thinking.

**Contextual Focus: shifting between divergent and convergent modes of thought.** Beyond the cognitive processes themselves it is also important to consider the cognitive flexibilities associated with them. Three types of cognitive flexibilities are related to divergent and convergent thinking: Spontaneous Flexibility, Reactive
Flexibility, and Contextual Focus. Contextual Focus, or the capacity to shift between convergent and divergent thinking, will be operationalized in the proposed research; But before describing Contextual Focus let us briefly describe the two other flexibilities mentioned in the context of their thinking processes. Spontaneous flexibility is defined as the ability to shift between different semantic or other domains to produce original ideas (Eslinger & Grattan, 1993). Spontaneous flexibility can be measured using the Alternate Usage Task (AUT) (Ionescu, 2012), by measuring how many different ideas or domains were used by the participant. The AUT consists in asking the participants to generate non-conventional (i.e. "alternative") original usages for common objects (e.g. Cup). The overall originality of the response should reflect spontaneous flexibility, the capacity to use multiple domains when generating ideas. Latent Semantic Analysis (LSA) can also be used to assess spontaneous flexibility; LSA consists of measuring the semantic distance between words and ideas. For example, a "cat" and a "lion" shall be closer semantically than a "lion" and a "bath tub." LSA is used to measure the semantic distance between all the ideas proposed by the participants during a divergent thinking task. LSA of ideas generated was found to be correlated with originality ratings (Forster & Dunbar, 2009). In the context of scientific reasoning, spontaneous flexibility, along with the process of divergent thinking, is the mental tool that will allow us to be creative when generating variables or hypotheses, but also sometimes when interpreting results of experiments.
Reactive flexibility is defined as the ability to shift between mental sets (Eslinger & Grattan, 1993). Changing strategy after receiving negative feedback would be a good example of reactive flexibility. Reactive flexibility can be measured using the Wisconsin Card Sorting Task (Tomer, Fisher, Giladi & Aharon-Peretz, 2002). The Wisconsin Card Sorting Task (WCST) is traditionally used by neurologists to measure frontal lobe dysfunction. The task itself is quite simple. The participant is presented with four decks of cards and individual stimulus cards that they have to sort into one of the four sets. Each stimulus card is composed of a certain number of colored shapes. At this point, the participant does not know what the "correct" rule, determined by the experimenter, is. The stimulus card could be associated with deck 1 (i.e., shape: triangle), deck 2 (i.e., color: yellow), or deck 4 (number: three). After a few rounds, the participant will find the correct strategy (shape, color, or number) to classify the stimulus card, but after a certain number of further rounds, the rule changes (with no explicit mention of the rule change). So, after several rounds using the same sorting strategy (e.g., shape), the rule changes (e.g., color) to force the participant to adapt. Subsequently using the "old" strategy, the participants will receive negative feedback, and then, will have to modify their strategy to discover the new correct sorting one. This change of strategy after receiving negative feedback embodies reactive flexibility. When a participant fails to switch to a new strategy, while re-using the same old obsolete one, this type of mistake is referred to as "perseverative error." Reactive flexibility is measured by counting the number of perseverative errors made by the participant—how many
times the participant fails to adapt his strategy after receiving negative feedback. In
the context of scientific reasoning, reactive flexibility could be related to its
evaluative process. After receiving results from experiments, some variables or
operators might have been found irrelevant and therefore should not be considered
any more when explaining a scientific phenomenon.

While spontaneous and reactive flexibility are taking place within divergent or
convergent thinking mental search, Contextual Focus is defined as the shifting between
divergent and convergent thinking. We constantly switch between convergent and
divergent thinking because complex reasoning tasks involve multiple goal-directed tasks
between the current and the goal state, such as reformulating the initial problem, retuning
the goal state, and generating new operators to reach the newly defined goal state
(Chrysikou, 2006). Contextual Focus is a term that was coined by Gabora (2010) to
describe such a shift. “Fully cognitive modernity following the appearance of anatomical
modernity after 200,000 BP, was made possible by the onset of Contextual Focus (CF):
the ability to shift between an explicit convergent mode conducive to logic and refinement
of ideas, and an implicit divergent mode conducive to free-association, viewing situations
from radically new perspectives, concept combination, analogical thinking, and insight”
(Gabora & Smith, 2018). Some scientists even believe that this capacity to shift between
a more associational and analytical mode of thinking is related to the birth of creativity
and is one of the pillars of our evolution as a species (Gabora & Dipaola, 2012). The
notion of cognitive flexibility is related to our faculty for adapting to different situations.
"Being flexible" is often a substitute for "being able to adapt." Cognitive flexibility is
rooted in agile thinking, creative thinking, or any complex reasoning like scientific reasoning. Divergent thinking is related to associational reasoning; while convergent thinking is related to analytical reasoning. Spontaneous flexibility is related to our capacity to be flexible during divergent thinking; and, reactive flexibility allows us to change our way when we detect a discrepancy during convergent thinking. Both modes of thinking are continuously used during complex tasks. When solving a problem, during scientific reasoning, we constantly shift between those two modes of thinking, referred to as Contextual Focus (Gabora, 2010). The creativity literature is one of the first one, to our knowledge to look more closely at the importance of the switching back and forth between convergent and divergent thinking. A study investigating design thinking through think aloud protocols, highlighted the importance of this switch between divergent and convergent thinking (i.e Contextual Focus) (Goldschmidt, 2016). Beyond the fact that convergent and divergent thinking are important during creative thinking, the shift between the two modes of thinking is presented, along with empirical data, in Goldschmidt’s paper as potentially the key to creative thinking.

The cognitive neuroscience literature, and more especially brain network activation, studies posit that the Salience Network is associated with Contextual Focus, the shift between convergent (i.e. executive control) and divergent thinking (i.e. default-mode network; Koutstaal & Binks, 2015). When salience information is detected and integrated from the outside environment or mentally generated, the salience network is active. It is hypothesized that the salience network is responsible for the switch between the Executive Control Network and the Default-Mode Network (Uddin, Supekar, Ryali & Menon, 2011)
when something triggers hierarchical control (Sridharan, Levitin & Menon, 2008) and modulates cognitive systems (Eckert, Menon, Walczak, Ahlstrom, Denslow, Horwitz & Dubno, 2009). The brain regions of interest during the activation of the salience network are the anterior insula cortex (integrating sensory information with other information about how we feel) and the anterior cingulate cortex (monitoring and evaluating our progress).

The “Z” model integrates each of these types of flexibility as well as the cognitive mechanisms, also referred to as modes of thought, of convergent and divergent thinking. Divergent and convergent thinking are a more specific representation of the duet between exploration-exploitation, which in turn can be perceived as the cognitive needs created by the perception-action cycle. In the context of scientific reasoning, we should expect Contextual Focus to play a central role.

**Personality and Thinking**

This part of the literature review concerns only the secondary question related to the broader research presented in this manuscript. Cognitive mechanisms, and more broadly cognition, are dependent upon different factors. Some different personality types will be more related to better performance on convergent or divergent thinking tasks. Our beliefs about the malleability of our creativity or intelligence will also alter our reasoning performance.

Our ability to perform convergent and divergent thinking tasks is also related to who we are, the formative experiences we had, also, define us as thinkers. The Five-Factor Model is composed of Openness/Intellect, Neuroticism, Agreeableness,
Conscientiousness, and Extraversion. It is understood that Openness/Intellect are associated with a more cognitive exploration mode; nonetheless, they can be meaningfully separated: Openness and Intellect (DeYoung, Quilty, Peterson & Gray, 2014). Each sub-part of the Openness/Intellect factor shall be therefore defined as follows (DeYoung, Grazioplene, & Peterson, 2012): Openness is related to our reasoning process, using our perception, fantasy, emotions, and aesthetics, while intellect is related to our reasoning process, using abstract and semantic information.

Openness is connected to the bottom-up process that allows us to make connections and find patterns based on sensory experience (DeYoung, 2014), more related to a divergent thinking mode. Intellect is connected to a more top-down process that helps us analyze patterns in a logical way (DeYoung, 2014), more related to convergent thinking modes of thought. Prior studies have found that fluid intelligence (convergent thinking task) was more strongly associated with Intellect, rather than Openness.

In a study investigating the relationship between the big five personality dimensions and achievements in the arts and sciences, the factor of openness/intellect was more particularly scrutinized (Kaufmann et al., 2016). While the Openness/Intellect factor predicts achievements in school and work, social behavior, and physical and mental health (Benet-Martinez, 2006 in Kaufman et al. 2016); Kaufman and colleagues posit that this factor may also be divisible, in terms of its relations and predictions with other achievements such as the Arts and Sciences. After testing four different samples of the population, for obvious ecological validity purposes, each participant completed the Creative Achievement Questionnaire (CAQ), allowing the researcher to assess participant's
creativity in 10 different domains. The creative achievement in the sciences was computed using the scores of inventions and scientific discovery domains; while the creative achievements in the arts was computed using the scores of visual arts, music, dance, creative writing, humor, theater, and film. Kaufmann and colleagues found that Openness was more associated to creative achievements in the arts, whereas Intellect was more related to creative achievement in the sciences (2016). It is indubitable that achievements in the arts or the sciences, both require, convergent and divergent thinking. However, it is also evident that Sciences and Arts may rely differently on convergent or divergent thinking. Therefore, finding that Openness and Intellect would predict different results for creative achievements in Arts or Sciences, emphasizes the importance of considering the role of personality during reasoning and cognition. We argue here that people who score higher in the intellect dimension shall display an advantage at scientific reasoning. In addition, participants who score high in openness shall also perform better in the generation of variables (problem space search) during scientific reasoning task; and participants who score high in the intellect dimension shall display an advantage during the evaluative search of the scientific reasoning process.

**Model Z: An Integrated Perspective to Scientific Reasoning**

For the purpose of clarity, we propose first to consider the scientific reasoning process from a broader perspective: a two-step process consisting of contemplating a divergent explorative state of mind before pursuing a convergent exploitative goal. When zooming out, the scientific reasoning process can be reduced to an explorative phase.
preceding an exploitative one, comparable to the perception-action cycle. Please note that hypothesis and experimental space search are included within the two larger two steps, problem space exploration and evaluative space exploitation (Figure 9). They are both comprised in one of the two broader categories. First, the hypothesis space search outcome belongs to the more divergent explorative state of mind (i.e. problem space exploration); while secondly, the experimental space search outcome belongs to the more convergent exploitative goal (i.e. evaluative space exploitation). In sum, this broader categorization of scientific reasoning is based on the overall state of mind and goal of the thinker. This more abstract categorization will help us to break down and introduce the “Z” model in this next section. However, please note that during the hypothesis and experimental space search, while the overall goal is first divergent then convergent in nature, each of those steps requires both modes of thinking. Because this research only targets the broader version of the “Z” model, the hypothesis and experimental space search will be discussed in more length during the discussion of this dissertation. This current research focuses on the broader searches in the problem and evaluative spaces.
The “Z” model (Figure 10), as previously stated, is a descriptive model that integrates previous findings on cognition. Findings that were scrutinized and categorized as suggested by the iCASA thinking framework (Koutstaal, 2012). The “Z” model does not only combine findings about specific cognitive mechanisms at play during scientific reasoning, but it also illustrates this more abstract, step-by-step, model “Alpha” of scientific reasoning. The “Z” model is situated at the heart of the perception-action cycle. The “Z” model holistically incorporates what we know about the different forms of mental search in scientific reasoning; but it also, in a more atomistic fashion, depicts the cognitive mechanisms and their need for flexible exploration and exploitation.
According to the “Alpha” model, scientific reasoning can be represented by a succession of searches in different spaces. In order to give a holistic view of the “Z” model to the reader, I will first break down the “Z” model into the two broader steps of scientific reasoning. The goal in this section about the “Z” model is to 1- describe the model from a scientific reasoning perspective (i.e. “Alpha” Model), then 2- bridge the model with the cognitive mechanisms (i.e. convergent thinking, divergent thinking, and Contextual Focus).

**Problem space exploration.** As reasoning is triggered by a question, scientific reasoning is prompted by a problem or a question about a puzzling phenomenon (top right corner of the model). During the initial searching phase, as for any other type of problems, the solver starts by exploring the problem space (Newell & Simon, 1972). The problem is
represented by the gap between a current state and the hoped-for goal state. The aim of the
problem solver is then to mentally generate and use operators in a goal-directed sequence
to reach the goal state (Chrysikou, 2006). The first goal after formulating a question or a
problem will be to search for variables/operators that could be used to explain and better
understand the phenomenon, the question at stake. The physical shape of the model to
represent the problem space exploration is represented by an "open" angle, symbolizing
the exploration, the divergent thinking search (Figure 11). The problem space is explored
via internal search, in the mind, using our knowledge about the phenomenon, and previous
hypotheses or theories. As represented, the mode of choice for the exploration of the
problem space is divergent thinking, which could be translated into the activation of the
default mode network. Spontaneous flexibility, associated with generative divergent
thinking search, will facilitate a deeper search for operators and variables. We will move
from surface features of the problem toward more in-depth features that will facilitate
stronger associations. As suggested by the serial order effect, the default-mode network
activation will allow us to find more connections, find new variables and operators that we
may not have considered initially resulting in the expansion of the problem space.
Concurrently, we engage in the hypothesis space search. This is one of the reasons why
hypothesis space search is comprised within the problem space exploration. The state of
mind during the exploration of the problem space could be described overall as bottom-up
oriented, reconstructing and reducing mentally the gap between the current state and the
goal state, using associative thinking such as, for example, analogical reasoning. I will
elaborate more on the importance of analogical reasoning in the section dedicated to hypothesis space search (see discussion section).

![Diagram](image)

*Figure 11. Problem space exploration in the “Z” model of scientific reasoning*

**Problem space exploration and insight problem-solving.** The idea landscape described by Koutstaal in *The Agile Mind* (2012) where ideas go in and out of consciousness is critical here to understand the problem space exploration. I argue here that we never stop exploring the problem space as we progress through the different steps. Even if our attentional control is geared toward the problem space at the beginning and as we move forward with our reasoning, we still continue to unconsciously explore the problem space (in the model “Alpha,” this constant problem exploration search is represented by a green line). When suddenly a variable of importance comes to mind at any point during the reasoning process, we integrate this finding, for example, to revise our problem state (e.g. Did I ask the right question?), or the hypothesis space search (e.g. I did not think about this important variable, I need to revise my hypothesis), or again the
evaluative space (e.g. this important variable would have helped me better interpret my results). Sometimes, in an "ah-ha" moment, we recognize instantly that the emerging variable is key to solving the problem. One of the most famous stories of insight is probably from Archimedes who, after days of struggle to find a way to measure the volume of the King's gold crown, was struck by the answer while in his bath: "EUREKA!" The main characteristic of insight problems is that when discovering the problem, the solver often misinterprets it and therefore focuses on the wrong operators, which will lead him to an "impasse," he will then need to restructure the problem to solve it (Fleck, 2008). Insight problems commonly involve the solver failing to explore the problem space correctly (e.g. John traveled the world without spending a single cent. Why?). Recent literature posits that insight problems are composed of two phases. The first step is the initial searching phase when the solver utilizes wrong operators (e.g. why someone can travel around the world for free? John is a pilot - John's mother owns the airplane company). Then the restructuring phase, or incubation phase, starts after the solver reaches the "impasse" (Lv, 2015). The impasse, corresponding to the current state not matching the goal state, is achieved when the solver is unable to generate additional potential solutions focusing on the wrong operators or variables. The insight problem answer habitually emerges to the mind of the solver in a magic "A-ha" moment after the restructuring phase (e.g. John is an animal!). The variable of importance emerges to consciousness and is readily recognized as the correct answer. The same way answers emerge to mind during insight problem solving, relevant variables or operators can also come to consciousness during scientific reasoning. Thus, such ideas may occur at any step of the scientific reasoning process because we are
constantly exploring, consciously or unconsciously, the problem space. Such phenomenon is illustrated by the internal mental nature of the exploration of the problem space as represented by the “Z” model. Because of the bottom-up role, related to perception, during this specific mental search of the scientific reasoning phase, the “Z” model suggests that divergent thinking plays here a more important role than convergent thinking. The experimental work presented here aims at investigating this part of the “Z” model.

**Evaluative space exploitation.** The evaluative space (Figure 12) corresponds to the honing phase when we capture the solution to the problem. We go through the experimental space search to find tools to test our hypothesis, with the objective to isolate one optimal solution to our problem. This type of search relies on the external world. This time we do not only mentally assess the viability of our associations, but we also concretely test them in the outside world before evaluating the outcome. We apply our knowledge of the environment to assess physically if the gap between the current state and goal state is reduced: Did the experimental space search lead to an acceptable answer? This top-down process corresponds to the action in the perception-action cycle, as described by Fuster (2015). The convergent thinking mode is the mode of choice: we focus; we make sure that every detail is taken care of; we exploit to the maximum our observations to extract the most useful information. The physical illustration of the evaluative space exploitation is represented by a closed angle that symbolizes this narrowing process. Reactive flexibility helps us sharpen the scientific process, by focusing on mistakes or ineffective strategies, for example. Because of the top-down role, related to action, during this specific mental search of the scientific reasoning phase, the “Z” model suggests that convergent thinking
plays here a more important role than divergent thinking. The experimental work presented here also aims at investigating this part of the “Z.” In addition, the shift between exploration and exploitation is omnipresent. As we, for example, detect that a strategy is not effective, we shift to the exploration mode to generate a new possible strategy and so on and so forth. For this reason, Contextual Focus should be considered as being central during the scientific reasoning process.

Figure 12. Evaluative space exploitation in the “Z” model of scientific reasoning

Cognitive flexibility: Shift between exploration of the problem space and exploitation of the evaluative space. Contextual Focus is the shift from the default network to the executive central network (represented in the model by the diagonal bar of the Z). It is crucial to understand the constant shift between divergent thinking – when we explore the problem space – and convergent thinking, when we exploit the evaluative space. The salience network, shifting between default mode and executive central network, is the brain activation pattern when engaging in Contextual Focus. Two pathways are
suspected to be linked with the shift between more associational problem space exploration and more analytical evaluative space exploitation. Firstly, driven by our perception and implicit associations, new operators and variables can emerge into our consciousness (i.e. idea landscape) in a bottom-up fashion, and our salience network will then recognize that this "new" information needs to be applied to our current goal state. Consequently, we shift and apply top-down, convergent thinking to reformulate, revise, and adapt our previous hypothesis, experimental design, or evaluation of results accordingly. So, we shift to a more convergent mode after being triggered by comparatively bottom up processes, such as insight problems, emerging variables or operators. Secondly, the trigger of this shift between an associational and analytical mode can also be top-down oriented. As we explore the problem, we consciously generate and regroup our knowledge toward the explanation of the problem or puzzling phenomenon. Besides, while evaluating our experimental results or observations, we may not be completely satisfied (salience network activation) and shift back to the problem space exploration to interpret our findings. Thus, we deliberately shift to a more divergent mode after being triggered by some more top down motivations during the evaluative space search.

It is not clear in the literature how the shift operates between analytical and associational mode. Nonetheless, Contextual Focus seems to be of great importance to understand scientific reasoning, and more generally thinking. The two perceptive paths that I described (i.e. bottom-up vs. top-down) are well captured in the cognitive neuroscience literature of prospective memory and more specifically the brain activation associated with prospective memory (PM). PM could be explained as remembering to remember, and so is
very close to the construct of Updating. PM is associated with an intention to do something at a given point in the future. A typical example of PM failure is the reusable grocery bag forgotten at home. There are two competing hypotheses regarding PM. The first hypothesis claims that prospective memory is driven by a top-down cognitive process; while the second hypothesis posits that prospective memory is dictated by a bottom-up type of cognitive process. Two specific cognitive processes have been related to PM (McDaniel, LaMontagne, Beck, Scullin & Braver, 2013). The first one is related to a top-down process, while the second one is related to a bottom-up process. The two processes leading to different strategies are considered and investigated by McDaniel and colleagues (2013). On the one hand, the top-down process of PM is presented as a tool to "maintain activation of the intention" while performing daily routines and tasks (Burgess, Quayle, & Frith, 2001). The top-down process is used to activate and extract the intention from the periodic memory (Craik, 1986). When individuals are successful in PM the top-down processes constantly reinforce the monitoring for the target event that is linked to the intention (Smith, 2003). For example, individuals could purposefully remind themselves every 5 minutes to take their reusable grocery shopping bag before leaving home. On the other hand, the bottom-up process of PM is associated with a multi-process theory (McDaniel & Einstein, 2000): it can be defined as "a transient process that is triggered by stimulus cues with strong associations to the PM intention" (e.g. individuals could purposefully associate the front door of their home with their grocery shopping bag). Findings suggest the existence of two neural pathways during PM. McDaniel and colleagues conducted an elegant experiment and showed that the same behavior of PM could be triggered by two
different strategies (2013). In fact, the brain imaging analysis lead to the following findings: the top-down process linked to PM is a more sustained brain activity and linked to regions of the brain associated with attentional monitoring. The bottom-up process linked to PM is related to a more transient activity of brain regions associated with stimulus detection. Finally, the difference in connections between the prefrontal cortex and other brain regions, when detecting the two kinds of stimulus (i.e. focal vs. non-focal) while engaged in PM, suggests that our brain elicits two different brain pathways during the same apparent behavior (PM).

During a top-down cognitive process, we have to constantly monitor and detect the appropriate stimulus to succeed in PM. During a bottom-up cognitive process, we only have to recognize a certain stimulus to succeed in PM. The study managed to unveil convincing brain imaging data concerning the existence of two different strategies during PM. The same phenomenon may happen during scientific reasoning and could explain the Contextual Focus shift between exploration to exploitation, and vice versa. Such a finding leads us to generate the following hypothesis: A Contextual Focus shift between problem space exploration and evaluative space exploitation can be triggered by a top-down, focused convergent thinking cognitive mode; and it can also be triggered by a bottom-up, defocused, divergent thinking mode of thought. It is clear that more studies should be performed to make a firmer connection between the PM dissociable neural routes and the bottom-up vs. top-down trigger of the Contextual Focus shift. Nonetheless, understanding the mechanisms of Contextual Focus will be important to understand better why our reasoning fails, or, alternatively, how we manage to adapt.
Some other recent findings suggest that during creative thinking the executive central network and the default mode happen to be activated at the same time (Beaty, Benedek, Silvia & Schacter, 2016). The collaborative nature of the two brain activation patterns, during creative thinking, suggests that the salience network, and therefore Contextual Focus, may not play a binary "switch," moderating, role. In addition to assessing the need for exploration-exploitation, Contextual Focus may also take into account the need for control (i.e. from automatic to effortful) within each type of search, thereby, mediating appropriate cognitive resources to convergent or divergent thinking searches. Considering the locus of control, as suggested by the iCASA thinking framework, may help understand in more fine grain the need and subsequent brain activation related to Contextual Focus. One may hypothesize that the salience network may play an attentional regulator role, not binary switching from exploration-exploitation (and vice versa), but rather allocate attentional control to each need (elaboration vs. exploitation) depending on their attentional needs. Some well-known behavior will be executed more automatically and therefore need less attentional control, while a more difficult or novel task will require more attention (top-down focus). The intriguing findings, stating the vanishing of the serial order effect when participants' fluid intelligence was controlled for (Beaty & Silvia, 2012), could be interpreted by considering that some individuals, with higher fluid intelligence – related to EF and executive control network activation – have a higher capacity for Contextual Focus. In other words, their ability for Contextual Focus – shifting between convergent and divergent thinking – allowed them to devote appropriate attentional effortful control to both (e.g. Divergent thinking 60% and Convergent thinking 40% of
attentional control). Given the central role of Contextual Focus, as depicted by the “Z”
model, related to switching between problem and evaluative space during scientific
reasoning, the role of Contextual Focus is also investigated.
Chapter 3: Research Questions

The goal of this experiment was to test the new model of scientific reasoning (i.e. “Z” model) and thus investigate the role of the different cognitive mechanisms of interest (i.e. divergent and convergent thinking) during scientific reasoning. In addition to testing the effect of the cognitive mechanisms involved in problem and evaluative space search during scientific reasoning, the “Z” model highlights the importance of Contextual Focus during scientific reasoning. Therefore, Contextual Focus was tested to investigate its predictive power in relation to performance on our specific measure of scientific reasoning (the Bouncing Ball Reasoning Task; BBRT) and also in relation to performance on a broader measure of scientific reasoning (the Lawson Test of Scientific Reasoning; LTSR). In addition, we tested the predictive power of Intellect and Openness over scientific reasoning performances.

Concerning the exploration of the problem space during scientific reasoning, a main question was tested: Does any type of cognitive experimental priming lead to an advantage when exploring the problem space? We hypothesized that participants experimentally primed to think divergently should perform better during the exploration of the problem space during a scientific reasoning task. Specifically, we hypothesized that:

1. Participants in the divergent thinking group should generate more hypotheses than the two other groups.
2. Participants in the divergent thinking group should generate more categories of variables than the two other groups.

3. Participants in the divergent thinking group should switch between categories of variables more often than the two other groups.

Concerning the exploitation of the evaluative space during scientific reasoning, the main question was tested was whether any type of cognitive experimental priming lead to an advantage when exploiting the problem space? We hypothesized that participants experimentally primed to think convergently should perform better during the exploitation of the evaluative space during a scientific reasoning task. Specifically, we hypothesized that:

4. Participants in the convergent thinking group should display fewer categorical errors after receiving problem-constraining information than the two other groups.

5. Participants in the convergent thinking group should generate explanations leading to more potential points, indicating engagement in more appropriately constrained thinking than the two other groups.

6. Participants in the convergent thinking group should provide more accurate and complete final answers/explanations than the two other groups.

7. Participants in the convergent thinking group should show higher correspondence between the actual accuracy of their explanations and their confidence in their answers in regard to the goal state/solution in comparison with the two other groups.
Concerning the role of Contextual Focus and personality traits (i.e. openness and intellect) during scientific reasoning, a main question was to be tested: Can Contextual Focus, Intellect and Openness scores predict performance on scientific reasoning tasks? Two tasks of scientific reasoning and three performance measures were considered. First, using a specific scientific reasoning problem (BBRT) exploration of problem space and exploitation of evaluative space performances were predicted using Contextual Focus, Intellect and Openness scores. Secondly, performance on a broader measure of scientific reasoning was predicted using the same variables, after controlling for performance on the more specific reasoning scores (BBRT). We hypothesized that:

8. Ability during exploration of the problem space should be significantly predicted by Contextual Focus and Openness.

9. Ability during exploitation of the evaluative space should be significantly predicted by Contextual Focus and Intellect.

10. Broader score of scientific reasoning should be significantly predicted by Contextual Focus, Intellect and Openness.
Chapter 4: Rationale for Operationalizations and Modifications

In this part we will describe the operational definitions used for our research. Some measures/tasks have been modified to better match the intended constructs (e.g. Remote Associate Task → Convergent thinking mode); therefore, we realized that it would enhance clarity to introduce them separately from the method section. While this part aims to answer “What are the chosen operational definitions” to match the constructs of interest, the reader will be informed “How they were used” during our research in the method section. But, before describing the measures and priming tasks, we chose to provide a rationale for choosing a think aloud protocol to collect the main data of interest: scientific reasoning.

Concerning the process for collecting our main dependent variable, the scientific reasoning (BBRT), a think aloud protocol was preferred rather than a traditional behavioral/written protocol. A think aloud protocol has the advantage of capturing the cognitive processes and ideas that people generate as they actually work on a task, from their short-term and working memory, rather than their long-term memory (Ericsson & Simon, 1980, 1993). We chose a think aloud protocol for two main reasons. The first reason is related to the nature of the data we wanted to collect: beyond the overall behavioral results of the participants, we are also interested in the trajectory of their ideas over time. The second reason is experimental by nature. Since we prime participants between the scientific reasoning tasks to think convergently, divergently, or neither (control group), we
suspected that collecting written answers would involuntarily prime a convergent processing mode in our participants. For example, some participants might exhibit a more top-down focus while providing the answers to the scientific reasoning task (e.g. searching in memory for the correct spelling of a particular word, trying to organize their thoughts before writing them down, or focusing on the legibility of their handwriting). Therefore, a think aloud protocol seemed more appropriate to allow us to collect the type of data we needed while minimizing interference with the intended cognitive manipulations of procedurally priming either convergent or divergent thinking.

**Experimental Priming Tasks: Convergent and Divergent Thinking**

Because it is difficult, some may argue impossible, to find a task that is only convergent or divergent, we modified slightly some existing convergent and divergent thinking tasks to insure a more efficient experimental priming.

For the convergent thinking priming task, we chose to use the Remote Associate Task (RAT). The original RAT was developed by Mednick (1962) and requires participants to find a single target word that can be associated with three stimulus words (e.g. Cottage, Swiss, Cake □ Cheese). RAT problems can traditionally be employed using two different types of strategies, insight vs. non-insight (Cranford & Moss, 2012). RAT can be solved through problem space exploration (Newell, 1990) when the participants generate and test different target words until a "hit" is found. Alternatively, RAT can also be solved through insight, wherein the target word suddenly emerges into mind followed by an "A-HA" moment (Cranford & Moss, 2012). While the first strategy described is top-down oriented,
requiring the desired convergent thinking mode of thinking, the second strategy describes a more bottom-up oriented process, closer to a divergent thinking cognitive mode. Therefore, in order to reduce the potentiality of triggering a divergent thinking mode and to more strongly and selectively emphasize the convergent thinking aspect of the task, we slightly modified the task from the original. We displayed three hyphens (i.e., “---”) in front or before each of the stimulus words to indicate the position of the target word (e.g. Cottage---, Swiss---, ---Cake). This modification aims to reduce the need to think divergently, because the participant does not have to think as flexibly to find out where the target word should be positioned. In addition, given that there were two priming phases, the specific items selected for the two convergent thinking priming tasks (12 RAT puzzles each) were chosen to be comparable based on their overall difficulty and average time to be solved as described by a normative study (Bowden & Jung-Beeman, 2003).

For the divergent thinking group, we chose to use the Alternate Uses Test (AUT) to prime this particular mode of thinking. The AUT consists in asking the participants to generate as many original possible uses for a common object (e.g. "brick," "newspaper"). Studies have posited a positive relationship between this task and real-life measures of creative achievement (e.g. gaining patents, producing novels and plays, the creation of professional organizations and businesses) (Torrance, 1988; Plucker, 1999). Traditionally, the number of original ideas generated by a participant is used as the dependent variable. We also modified this task from the original because, during the task, some participants may find themselves inactive because of a lack of ideas. To encourage them to remain in the divergent thinking mode throughout the full priming phase of five (5) minutes, we
offered them two common objects instead of only one. Participants were instructed that they could generate ideas back and forth from one word to another as they wished. Priming with the AUT task has been shown to facilitate subsequent performance on unrelated verbal and nonverbal insight problem-solving tasks (Chrysikou, 2006; Wen, Butler, & Koutstaal, 2013), suggesting that engagement in this divergent thinking task can enhance other forms of open-ended exploratory search. Conversely, priming with a version of the RAT task has previously been used to induce a convergent thinking mode, in which task performance is facilitated by a more local and narrowly-focused mode of task processing (Colzato, van den Wildenberg, & Hommel, 2013).

Our goal for the control group was to find a control task that would neither induce convergent nor divergent thinking. The Word Association (WA) has been found to prompt participants to engage in a very automatic mode of thinking, distinct from convergent and divergent thinking, and to yield behavioral results highly comparable to a no-task group (Wen, Butler, & Koutstaal, 2013). The WA task consists in asking the participant to write the first word that comes to mind after reading a given stimulus word (e.g. “Dog” → “Walk”). We did not modify this task from the original, and like for the two previous tasks, the WA was scored (total number of words generated).

**Primary Dependent Measure: Scientific Reasoning**

Our primary scientific reasoning task presented participants with a puzzling phenomenon: The Bouncing Balls Reasoning Task (BBRT). Two bouncing balls exhibited conflicting behaviors: one of the balls bounced higher than the other but completely
stopped after fewer bounces than the other one. Participants were asked to answer two questions regarding this puzzling phenomenon (Phase 1 and Phase 2 – see Appendix A). Participants were primed before each phase of the scientific reasoning task, with participants primed in the same thinking mode before Phase 1 and Phase 2 (that is, AUT-AUT, RAT-RAT, or WA-WA). Each participant’s think aloud data were recorded using the software Audacity via a headset microphone.

In Phase 1, participants were given five (5) minutes to generate hypotheses – that is, potential explanations – about this puzzling phenomenon. Participants’ answers were transcribed and coded. The focus of Phase 1 of our scientific reasoning task is the exploration of the problem space, and more particularly the generation of hypotheses. The overall number of hypotheses was coded as well as the categories to which they belonged. Given the nature of our scientific problem and previous data collected using this task, we found that the hypotheses generated could belong to four different categories:

- **Environment is different** (E). For example, one ball may have initially bounced on a harder surface such as asphalt and later bounced over a softer surface such as sand.

- **Balls are different** (B). For example, one ball may be a football, while the other is a basketball.

- **Actions on the Ball are different** (AB). For example, one ball was thrown differently with more strength than the other.
- *Psychological/Human Factor can explain the puzzling phenomenon* (PHF). For example, the person witnessing the phenomenon may have made a mistake in counting/evaluating the bounces.

In Phase 2, participants were given additional limiting or problem-constraining information about the phenomenon. Specifically, participants were told that, after running some experiments, the conclusion reveals that the balls were strictly identical and that the environment did not play a role. The participants were at this point in time invited to think aloud, during 5 minutes, about their final explanation(s). The same as was done during Phase 1, their answers were recorded, transcribed, and coded. Because of the Phase 2 probe, participants should now only focus on the *Action on the Ball* category. Therefore, the number of categorical errors (number of explanations belonging to a category that should have been discarded – E, B, PHF) was also coded along with the potential points awarded to their explanations. The explanations given during the second phase of the think aloud protocols were transcribed and awarded points as follows:

- 1 point per variable of interest mentioned (i.e. strength and angle of throw)
- 3 points if the interaction between the variables is mentioned or was alluded to but with little specificity (e.g. “the combination of the strength and angle of throw will explain the phenomenon”).
- 4 points if the interaction between the variables is mentioned specifically (e.g. “The red ball was thrown with not a lot of strength at a steeper angle, while the blue ball was thrown with more strength at a shallower angle”). The cumulative potential points generated by the explanations were coded. In
addition, at the end of the Phase 2, participants were asked to write down their final best explanation, and afterward they were also asked to indicate how confident they were that their explanation was correct (from 0 to 100 percent, with values above 50% indicating increasing confidence).

Secondary Dependent Measures

**Broader measure of scientific reasoning.** The Lawson Test of Scientific Reasoning (LTSR) is a 24-item multiple-choice test widely used in education (Lawson, 1978). The test was modified in 2000, and was used here. The LTSR measures different sub-skills of scientific reasoning including conservation of mass and volume, proportional reasoning, control of variables, probabilistic thinking, correlational thinking, and hypothetical-deductive reasoning. The number of questions for each of the subskills is shown in Table 1. Participants completed a Qualtrics computerized version of the LTSR.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Skill Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3, 4</td>
<td>Conservation of Mass and Volume</td>
</tr>
<tr>
<td>5, 6, 7, 8</td>
<td>Proportional Reasoning</td>
</tr>
<tr>
<td>9, 10, 11, 12, 13, 14</td>
<td>Control of Variables</td>
</tr>
<tr>
<td>15, 16, 17, 18</td>
<td>Probabilistic Thinking</td>
</tr>
<tr>
<td>19, 20</td>
<td>Correlational Thinking</td>
</tr>
<tr>
<td>21, 22, 23, 24</td>
<td>Hypothetical-deductive Reasoning</td>
</tr>
</tbody>
</table>
Personality Measure: Big Five Aspect Scales

The Big Five Aspect Scale (BFAS) (DeYoung, Quilty & Peterson, 2007), a specific measurement of the Big Five personality traits is one of the most commonly used in classifications of personality traits. It was conceived by investigating the correlation of specific personality trait factors rather than theoretical research about abstract constructs (John, Naumann, & Soto, 2008). The BFAS is a 100-item questionnaire that measures five psychological factors, which are divided themselves into two distinct facets, each comprised of 10 items:

- Openness to Experience (Openness/Intellect)
- Conscientiousness (Orderliness/Industriousness)
- Extraversion (Enthusiasm/Assertiveness)
- Agreeableness (Politeness/Compassion)
- Neuroticism (Withdrawal/Volatility)

Each facet has been found to have internal reliability exceeding .75 (DeYoung, Quilty & Peterson, 2007).

The Five-Factor Model is composed of Openness/Intellect, Neuroticism, Agreeableness, Conscientiousness, and Extraversion. It is understood that Openness/Intellect are associated with a more cognitive exploration mode; nonetheless, they could be meaningfully separated: Openness and Intellect (DeYoung, Quilty, Peterson & Gray, 2014). Each sub-part of the Openness/Intellect factor are therefore defined as follows (DeYoung, Grazioplene, & Peterson, 2012):
• Openness echoes our reasoning process, using our perception, fantasy, emotions, and aesthetics.

• Intellect echoes our reasoning process, using abstract and semantic information.

Participants completed a Qualtrics computerized version of the BFAS.

**Contextual Focus**

We have defined the construct of Contextual Focus as the shift between convergent and divergent thinking (Gabora, 2010). This construct has yet to be operationally defined to the best of our knowledge. In a previous study (Quillien, Anderson & Koutstaal 2015a; Quillien, Anderson & Koutstaal, 2015b), we operationally defined Contextual Focus as the sum of the convergent and divergent thinking capacity. Someone who is capable of intense convergent and divergent thinking should also display a higher level of Contextual Focus than someone who has a moderate capacity to think convergently and divergently (Figure 13). Since we have demonstrated statistical relationships by operationalizing Contextual Focus in this manner and since it has been associated with executive functions (i.e. shifting and inhibition), we decided to apply the same protocol for this current project. The Contextual Focus score being define as a composite score of the participant’s convergent and divergent thinking. Participants completed two measures of figural convergent thinking tasks, the Elaboration Reproduction Task (ERT), then two figural divergent thinking tasks, The Ruff Test of Figural Fluency (RFFT). We deliberately chose to start with the convergent thinking task (ERT) before the divergent thinking one (RFFT). We
suspected the divergent thinking task of a larger carry over effect over the following convergent thinking one. Scores for each perceptual cognitive task were averaged to obtain one single convergent thinking and one single divergent thinking score. Then Z scores were computed for each perceptual cognitive task. The overall Contextual Focus capacity was calculated according to the following formula:

\[ CF = Z_{\text{conv}} + Z_{\text{div}} \]

![Diagram](image)

*Figure 13. Need for Contextual Focus conditional to convergent and divergent thinking capacity (e.g. Comparison Individual A and B)*

The Elaborative Reproduction Task (ERT) requires the participant to precisely reproduce (copy without tracing) a very abstract figure, within a limited time of 3 minutes (Figure 14) (Please note that the figure only represents a fraction of the design to be reproduced). This task requires participants to focus, reapplying what is seen on the model, while being careful to not make mistakes. Like traditional conceptions of convergent thinking, some features are being compared and synthesized for a correct final answer.
After two different ERT tasks during three minutes, the number of total correct segments reproduced correctly, in comparison to the originals, would be scored and averaged between the two trials to obtain the individual's perceptual-visual capacity to think convergently.

*Figure 14. Elaborative Reproduction Task (partial representation)*

The Ruff Figural Fluency Test (RFFT), unlike the AUT, is not a test of verbal fluency, but a test of figural fluency (Ruff, Light & Evans, 1987). Participants are instructed to create as many designs as they can, using an imposed arrangement of dots. Participants would traditionally read the following instructions:

“In front of you are three squares, each containing five dots. Note that the arrangement of the five dots is always the same. I want you to connect two or more dots by always using straight lines. The purpose of the test is for you to make as many designs or patterns as possible, but each design has to be different in some way from all the others.”
Similarly to the ERT, participants completed two versions of the RFFT (Figure 15) on two occasions each lasting three minutes. The total number of original designs was averaged between the two trials to obtain the individual's perceptual-visual capacity to think divergently. The patterns of dot arrangements that we chose for our two tasks are designated as IV and V in Figure 15.

*Figure 15. RFFT Parts 1 through 5, illustrating different dot arrangements for the RFFT subtests.*
Chapter 5: Methods

Participants

Participants included 132 students from a large mid-western University (46 male, mean age: 19.2 years, 14.15 years of education). Extra course credit was offered in compensation for participation. Using a block randomization procedure 44 students were randomly assigned to the convergent thinking group, 45 to the divergent thinking group, and 43 to the control group. Two participants were excluded for the Phase 2 of the BBRT, because of missing data. Six participants were excluded for the Final Score analyses of the BBRT, because of missing data. One participant was excluded from the Contextual Focus Analysis because of missing data.

Design

The design of the experiment was a between subject design using a think aloud protocol. The experiment encompasses three different conditions. Each condition is defined by the type of cognitive priming (i.e. convergent thinking, divergent thinking, or control). Participants were primed twice, each before completing two phases of a scientific problem solving task.

The dependent measures collected from the think aloud protocol were designed to capture participants’ idea formation and strategies. To track participants’ idea formation and strategies across the trajectory of their thinking when solving a puzzling scientific
problem (BBRT), we used a think aloud protocol to collect their thoughts while solving the puzzling scientific problem. From the transcripts of the think aloud, we also measured behavioral data related to the number of hypotheses they were able to generate (Phase 1) and their ability to draw the best conclusion, avoiding errors, after receiving additional information that constrained or limited the evaluative space about the scientific problem task (Phase 2). Each measure was fully or partially coded by two separates coders bling form the hypothesis and conditions. The inter-rater agreement was computed in the form of correlation. When the agreement is above 70%, it is usually considered as acceptable inter-rater reliability (Osborne, 2008).

** Priming Tasks **

**Convergent thinking.** Convergent thinking is often semantically measured or experimentally primed using the Remote Associations Task (RAT). In our experiment, participants in the convergent thinking group had 5 minutes to complete to the best of their ability twelves (12) RAT puzzles. It is important to clarify that the number of correct answers by RAT priming phases were scored for control purposes. The motivation and understanding of the task or other confounding variables may decrease the active participation in the task, and consequently the effect of the priming phase. Data from participants scoring below three standard deviations from the mean would be removed from the data set.

**Divergent thinking.** Divergent thinking is often semantically measured or experimentally primed using the Alternative Usage Task (AUT). In our experiment,
participants in the divergent thinking group had 5 minutes during each of the two priming phases to write as many original uses for the two presented words. Like for the convergent thinking group (i.e. RAT), the AUT was scored (e.g. overall number of ideas) for control purposes. Data from participants scoring below three standard deviations from the mean would be removed from the data set.

**Control task.** The Word Association (WA) task was chosen as a control task. In our experiment, participants in the control group had also 5 minutes to complete the task composed of eighty words (80). Control group participants were instructed to generate words after reading the stimulus one, and in case they finished the given list, they were instructed to go back to the beginning of the list and to keep generating new words to keep them cognitively active. Using the same rationale as discussed in our experimental groups, data from participants performing below three standard deviations below the mean (i.e. number of words generated) would be excluded from the statistical analysis.

**Measures**

**Scientific reasoning: Bouncing Balls Reasoning Task.** Our specific scientific reasoning task is divided into two phases. The first phase of the BBRT asked the participants to generate hypotheses regarding the possible reasons that might make the puzzling phenomenon possible during 5 minutes. The second phase of the BBRT asked the participants to formulate an explanation about the reasons that may have made the puzzling phenomenon possible during 5 minutes, after receiving limiting information about the
puzzling phenomenon that ruled out certain broad types of variables as possible accounts of the observed phenomenon.

**Contextual Focus.** Contextual Focus was measured for every participant. Our measure of Contextual Focus is a composite of a perceptual-motor convergent thinking score (ERT) and a perceptual-motor divergent thinking score (RFFT) computed from two figurally-based tasks. Each participant completed during 3 minutes two ERT tasks. The overall ERT score was determined by averaging the two trials.

**Lawson Test of Scientific Reasoning.** Participants would perform the LTSR at the end of the experiment using a computerized version of the LTSR on Qualtrics.

**Big Five Aspect Scales.** Participants performed a computerized version of the BFAS using Qualtrics. This automatic task was performed before completing the general scientific reasoning task, LTSR, to help the depletion of any potential experimental priming left in participant’s cognition. Only the scores for Openness and Intellect were computed and included for analysis purpose.

**Procedure**

After consenting to the study, participants completed different types of tasks, including paper-and-pencil, think-aloud, and computerized. Each participant completed the study in a laboratory room at the University of Minnesota. Upon arrival, participants were randomly assigned into one of three groups (i.e. control, convergent and divergent thinking groups) using block randomization. The first two tasks to be completed by the participants were the ERT and the RFFT, to determine their Contextual Focus ability by
measuring their individual perceptual-motor capacity in convergent and divergent thinking tasks. After completing the two figural tasks, participants were primed accordingly to their designated experimental group for five minutes. After the first priming, Phase 1 of the scientific reasoning problem was administered. All participants were asked during this phase to think aloud, for five (5) minutes, about explanations about the described puzzling phenomenon. Again, prior to the second phase of the scientific reasoning task, participants were primed in accordance with their experimental group. After completing Phase 2 of the scientific reasoning task (i.e. exploitation of the evaluative space), participants completed the computerized personality questionnaire (i.e. BFAS) and finally the computerized multiple-choice test of scientific reasoning (LTSR).

After completing the last two tasks, participants were debriefed, compensated, and dismissed (Figure 16).
Figure 16. Research design by conditions and time
Chapter 6: Results

The primary goal of this experiment was to test the “Z” model of scientific reasoning by investigating the role of divergent and convergent thinking during scientific reasoning (corresponding especially to exploration of the problem space and exploitation of the evaluative space respectively). The roles of Contextual Focus, Intellect and Openness during scientific reasoning were also investigated.

The first group of research questions focused on the potential advantage of being primed in a divergent thinking mode during the exploration of the problem space (i.e. Phase 1 of the BBRT):

1. Participants in the divergent thinking group should generate more hypotheses than the two other groups.
2. Participants in the divergent thinking group should generate more categories of variables than the two other groups.
3. Participants in the divergent thinking group should switch between categories of variables more often than the two other groups.

The second group of research questions focused on the potential advantage of being primed in a convergent thinking mode during the exploitation of the evaluative space (i.e. Phase 2 of the BBRT):

4. Participants in the convergent thinking group should display fewer categorical errors after receiving constraining information than the two other groups.
5. Participants in the convergent thinking group should generate explanations leading to more potential points than the two other groups.

6. Participants in the convergent thinking group should provide a better final answer/explanation than the two other groups.

7. Participants in the convergent thinking group should better evaluate their explanation in regard to the goal state/solution than the two other groups.

The third group of research questions focused on the role of Contextual Focus, Intellect and Openness during scientific reasoning:

8. Ability during exploration of the problem space (i.e. Phase 1 of the BBRT) should be significantly predicted by Contextual Focus and Openness.

9. Ability during exploitation of the evaluative space (i.e. Phase 2 of the BBRT) should be significantly predicted by Contextual Focus and Intellect.

10. Higher scores on the broader test of scientific reasoning (the LTSR) should be significantly predicted by Contextual Focus, Intellect and Openness.

**Descriptive Statistics and Correlations**

Descriptive statistics, including means and standard deviations for all variables relevant to the 10 research questions, are presented in Table 2. Correlations between all variables are presented in Table 3.
Table 2

Descriptive Statistics for All Variables

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBRT Phase 1 Hypotheses</td>
<td>12.06</td>
<td>5.62</td>
<td>1 – 26</td>
</tr>
<tr>
<td>BBRT Phase 1 Categories</td>
<td>3.36</td>
<td>0.82</td>
<td>1 – 4</td>
</tr>
<tr>
<td>BBRT Phase 1 Switch Categories</td>
<td>5.88</td>
<td>3.43</td>
<td>0 – 19</td>
</tr>
<tr>
<td>BBRT Phase 1 Overall</td>
<td>0.00</td>
<td>0.87</td>
<td>-2.18 – 2.11</td>
</tr>
<tr>
<td>BBRT Phase 2 Errors</td>
<td>3.98</td>
<td>4.10</td>
<td>0 – 18</td>
</tr>
<tr>
<td>BBRT Phase 2 Points awarded</td>
<td>1.55</td>
<td>1.20</td>
<td>0 – 4</td>
</tr>
<tr>
<td>BBRT Phase 2 Final score (written)</td>
<td>0.79</td>
<td>0.74</td>
<td>0 – 4</td>
</tr>
<tr>
<td>BBRT Phase 2 Overall</td>
<td>0.01</td>
<td>0.86</td>
<td>-1.16 – 4.34</td>
</tr>
<tr>
<td>Contextual Focus</td>
<td>0.00</td>
<td>0.78</td>
<td>-1.87 – 1.63</td>
</tr>
<tr>
<td>LTSR</td>
<td>15.01</td>
<td>4.15</td>
<td>2 – 23</td>
</tr>
<tr>
<td>Intellect</td>
<td>35.88</td>
<td>5.33</td>
<td>20 – 47</td>
</tr>
<tr>
<td>Openness</td>
<td>37.42</td>
<td>6.23</td>
<td>22 – 50</td>
</tr>
</tbody>
</table>

Note: BBRT = Bouncing Ball Reasoning Task, BBRT Phase 1 Overall is the overall performance during the exploration of the problem space (Zhyp+Zcat+Zswitch); BBRT Phase 2 Overall is the overall performance during the exploitation of the evaluative space (Zawarded+Zfinalscore); LTSR = Lawson Test of Scientific Reasoning
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. P1 Score</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. P1 Hypotheses</td>
<td>.89**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. P1 Categories</td>
<td>.82**</td>
<td>.55**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. P1 Switching</td>
<td>.90**</td>
<td>.76**</td>
<td>.60**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. P2 Score</td>
<td>.31**</td>
<td>.28**</td>
<td>.28**</td>
<td>.25**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. P2 Points</td>
<td>.30**</td>
<td>.24**</td>
<td>.30**</td>
<td>.25**</td>
<td>.86**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. P2 Errors</td>
<td>.27**</td>
<td>.38**</td>
<td>.08</td>
<td>.26**</td>
<td>-.26**</td>
<td>-.29**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. P2 Final Score</td>
<td>.22*</td>
<td>.23*</td>
<td>.18*</td>
<td>.18*</td>
<td>.86**</td>
<td>.49**</td>
<td>-.16</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Context Focus</td>
<td>.34**</td>
<td>.30**</td>
<td>.37**</td>
<td>.22**</td>
<td>.14</td>
<td>.13</td>
<td>.10</td>
<td>.09</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Lawson Score</td>
<td>.33**</td>
<td>.29**</td>
<td>.30**</td>
<td>.28**</td>
<td>.16</td>
<td>.22*</td>
<td>.00</td>
<td>.05</td>
<td>.33**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>11. Intellect</td>
<td>.27**</td>
<td>.20*</td>
<td>.32**</td>
<td>.19*</td>
<td>.20*</td>
<td>.19*</td>
<td>-.13</td>
<td>.15</td>
<td>.19*</td>
<td>.34**</td>
<td>1</td>
</tr>
<tr>
<td>12. Openness</td>
<td>-.02</td>
<td>-.03</td>
<td>.05</td>
<td>-.07</td>
<td>-.05</td>
<td>-.12</td>
<td>-.02</td>
<td>.04</td>
<td>-.05</td>
<td>.13</td>
<td>.22*</td>
</tr>
</tbody>
</table>

*p < 0.05, **p < 0.01
Phase 1: Exploration of the Problem Space

Our first hypothesis regarding the exploration of the problem space (Phase 1 of the BBRT) was: participants in the divergent thinking group should generate more hypotheses than participants in the two other groups. Our second hypothesis was that participants in the divergent thinking group should generate more categories of variables than the two other groups. Our third hypothesis was that participants in the divergent thinking group should switch between categories of variables more often than the two other groups.

During the first phase of the Bouncing Ball Reasoning Task (BBRT), participants were invited to formulate, during a think aloud protocol, hypotheses explaining a puzzling phenomenon involving bouncing balls. Three variables were coded from the transcripts for Phase 1 by two coders blind to the research question and hypotheses. First the coder reported the number of hypotheses generated by the participants during the think-aloud. Inter-rater reliability, for this variable and the subsequent ones, was computed by correlating scores between two coders. For this variable, using 20 % of the data, the inter-rater reliability was good ($r = .94$). The second variable of interest was the total number of categories cited by the participants. The coders scored the number of categories reported by the participants. Inter-rater reliability between two coders for this variable, using 20% of the data, was good ($r = .82$). Finally, the number of switches between categories
was reported by the coders. Inter-rater reliability between two coders for this variable, using 20% of the data, was good ($r = .93$).

**Number of hypotheses generated.** The first research question was whether experimentally inducing a divergent mode of thinking would lead to an advantage in the generation of hypotheses during Phase 1. The dependent variable in this analysis, number of hypotheses generated, was tabulated from the coded think aloud data. An ANOVA was run with one between-subjects factor. The between-subjects factor, Condition had three levels: divergent prime (AUT), convergent prime (RAT), and no prime (CON). A main effect of Condition was found, $F(2, 129) = 7.74$, $p = .001$, $\eta^2_p = .11$. Post hoc tests using the Tukey HSD procedure revealed that participants, as shown in Figure 17, in the divergent thinking group ($M = 14.58, SD = 5.8$) generated more hypotheses on average than did participants in the convergent thinking group ($M = 10.41, SD = 4.34, p = .001$) or the control group ($M = 11.12, SD = 5.75, p = .01$), whereas participants in the latter two groups did not differ ($p = .81$). Our hypothesis for this first research question was corroborated; participants in the divergent thinking group generated more hypotheses than the two other groups.
Figure 17. Number of hypotheses generated during Phase 1 per Condition. The error bars represent 95% CIs.

To understand the effect of time and condition during the BBRT a 3x3 ANOVA was run with one between-subjects factor and one within-subjects factor (see descriptive statistics Table 4). The between-subjects factor Condition had three levels: divergent prime (AUT), convergent prime (RAT), and no prime (CON). The within-subjects factor Time had 3 levels formed by dividing the 300 seconds of Phase 1 into 3 successive intervals: 0-100 seconds (“100”), 100-200 seconds (“200”), and 200-300 seconds (“300”), within which the total number of hypotheses generated during each time interval was tabulated. A repeated measures ANOVA did not find any interaction of Time and Condition ($F(4, 258) = .42, p = .79, \eta^2_p = .007$). The main effect of Condition found previously was...
reproduced (see previous analysis). Finally, there was also a main effect of Time, $F(2, 258) = 68.69, p < .001, \eta^2_p = .35$. As Figure 18 shows, participants generated fewer ideas with time. This observation was confirmed by the significant linear contrast ($F(1, 129) = 114.54, p < .001, \eta^2_p = .47$), which indicates that the number of ideas per time period decreased over time. In addition, the significant quadratic contrast ($F(1, 129) = 12.91, p < .001, \eta^2_p = .09$), suggests that hypothesis generation may have been reaching an asymptote at the end of 300 seconds.

Table 4

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Descriptive Statistics for Number of Hypotheses by Time Intervals and by Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Interval</td>
<td>Condition</td>
</tr>
<tr>
<td>0-100 Seconds</td>
<td>Divergent Thinking</td>
</tr>
<tr>
<td></td>
<td>Control</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>100- 200 Seconds</td>
<td>Divergent Thinking</td>
</tr>
<tr>
<td></td>
<td>Control</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>200- 300 Seconds</td>
<td>Divergent Thinking</td>
</tr>
<tr>
<td></td>
<td>Control</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
</tbody>
</table>
Figure 18. Number of hypotheses generated during Phase 1, averaged over 100-second intervals. The error bars represent 95% CIs.

**Number of categories generated.** The second research question pertaining to the exploration of the problem space was whether experimentally inducing a divergent mode of thinking would lead to an advantage in the generation of different categories of hypotheses during Phase 1. The dependent variable in this analysis, number of categories generated, was tabulated from the coded think aloud data. The same analysis strategy was used as for the number of hypotheses generated. First, an ANOVA was run with one between-subjects factor. The between-subjects factor, Condition, had three levels: divergent prime (AUT), convergent prime (RAT), and no prime (CON). No significant result was found, $F(2, 129) = .58, p = .56, \eta^2_p = .01$. Our hypothesis for this second research question was not supported: participants in the divergent thinking group did not generate more categories than the two other groups. In addition, to understand the effect of time and
condition during the BBRT a 3x3 ANOVA was run with one between-subjects factor and one within-subjects factor (see descriptive statistics Table 5). The between-subjects factor Condition had three levels: divergent prime (AUT), convergent prime (RAT), and no prime (CON). The within-subjects factor Time had 3 levels formed by dividing the 300 seconds into 3 successive intervals: 0-100 seconds (“100”), 100-200 seconds (“200”), and 200-300 seconds (“300”), within which the total number of categories generated during each time interval was tabulated. The Mauchly’s sphericity test was significant, indicating that the assumption of sphericity was violated. Therefore, we used the Greenhouse-Geisser correction. As already reported above, there was no effect of Condition ($F(2, 129) = .58, p = .56, \eta^2_p = .01$). There was a main effect of Time, $F(1.55,199.43) = 236.82, p < .001, \eta^2_p = .65$. As Figure 19 shows, participants generated fewer categories with time. This observation was confirmed by the significant linear contrast ($F(1, 129) = 334.90, p < .001, \eta^2_p = .72$), which indicates that the number of categories decreased over time. In addition, the significant quadratic contrast ($F(1, 129) = 95.48, p < .001, \eta^2_p = .43$), suggests that categories generation may have been reaching an asymptote at the end of 300 seconds. Finally, there was no interaction of Condition and Time ($F(3.09, 199.43) = .31, p = .82, \eta^2_p = .01$).
Table 5
Descriptive Statistics for Number of Categories Generated by Time Intervals and by Conditions

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Condition</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-100 Seconds</td>
<td>Divergent Thinking</td>
<td>2.53</td>
<td>1.01</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>2.42</td>
<td>1.07</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
<td>2.50</td>
<td>1.02</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.48</td>
<td>1.03</td>
<td>132</td>
</tr>
<tr>
<td>100-200 Seconds</td>
<td>Divergent Thinking</td>
<td>0.56</td>
<td>0.72</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>0.49</td>
<td>0.55</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
<td>0.64</td>
<td>0.65</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.56</td>
<td>0.64</td>
<td>132</td>
</tr>
<tr>
<td>200-300 Seconds</td>
<td>Divergent Thinking</td>
<td>0.36</td>
<td>0.64</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>0.35</td>
<td>0.57</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
<td>0.23</td>
<td>0.42</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.31</td>
<td>0.55</td>
<td>132</td>
</tr>
</tbody>
</table>

Figure 19. Number of categories generated during Phase 1, averaged over 100-second intervals. The error bars represent 95% CIs.
**Number of switches between categories.** The third research question pertaining to the exploration of the problem space was whether experimentally inducing a divergent mode of thinking would lead to more switching between categories during Phase 1. The dependent variable in this analysis, number of switches between categories, was tabulated from the coded think aloud data. First, an ANOVA was run with one between-subjects factor. The between-subjects factor, Condition had three levels: divergent prime (AUT), convergent prime (RAT), and no prime (CON). No significant result was found, $F(2, 129) = 1.37, p = .26, \eta_p^2 = .02$. Our hypothesis for this third research question was not supported: participants in the divergent thinking group did not switch more often between categories than the two other groups.

To understand the effect of Time and Condition during the BBRT a 3x3 ANOVA was run with one between-subjects factor and one within-subjects factor (see descriptive statistics Table 6). The between-subjects factor Condition had three levels: divergent prime (AUT), convergent prime (RAT), and no prime (CON). The within-subjects factor Time had 3 levels formed by dividing the 300 seconds into 3 successive intervals: 0-100 seconds (“100”), 100-200 seconds (“200”), and 200-300 seconds (“300”), within which the total number of switches observed during each time interval was tabulated. A repeated measures ANOVA did not find any interaction of Time and Condition ($F(4, 258) = .61, p = .66, \eta_p^2 = .01$). Finally, there was a main effect of Time, $F(2, 258) = 9.90, p < .001, \eta_p^2 = .07$. As Figure 20 shows, participants exhibited fewer switches with time. This observation was confirmed by the significant linear contrast ($F(1, 129) = 17.51, p < .001, \eta_p^2 = .12$), which
indicates that the number of switches between categories decreased over time. In addition, no significant quadratic contrast was found ($F(1, 129) = 1.34, p = .25, \eta^2_p = .01$).

Table 6
Descriptive Statistics for Number of Switches Observed by Time Intervals and by Conditions

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Condition</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-100 Seconds</td>
<td>Divergent Thinking</td>
<td>2.47</td>
<td>1.96</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>2.30</td>
<td>1.67</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
<td>2.39</td>
<td>1.83</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.39</td>
<td>1.81</td>
<td>132</td>
</tr>
<tr>
<td>100- 200 Seconds</td>
<td>Divergent Thinking</td>
<td>2.04</td>
<td>1.76</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>1.49</td>
<td>1.35</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
<td>2.00</td>
<td>1.22</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1.85</td>
<td>1.47</td>
<td>132</td>
</tr>
<tr>
<td>200- 300 Seconds</td>
<td>Divergent Thinking</td>
<td>1.96</td>
<td>1.87</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>1.47</td>
<td>1.32</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
<td>1.50</td>
<td>1.37</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1.64</td>
<td>1.55</td>
<td>132</td>
</tr>
</tbody>
</table>
Figure 20. Number of switches between categories generated during Phase 1, averaged over 100-second intervals. The error bars represent 95% CIs.

**Phase 2: Exploitation of the Evaluative Space**

Our fourth hypothesis pertaining to the exploitation of the evaluative space (Phase 2 of the BBRT) was that participants in the convergent thinking group should display fewer categorical errors after receiving problem-constraining information than the two other groups. Our fifth hypothesis was that participants in the convergent thinking group should generate explanations leading to more potential awarded points than the two other groups. Our sixth hypothesis was that participants in the convergent thinking group should provide better final answers/explanations than the two other groups. Finally, our seventh hypothesis was that participants in the convergent thinking group should better evaluate their explanation in regard to the goal state/solution in comparison with the two other groups.
During the second phase of the Bouncing Ball Reasoning Task (BBRT), participants were invited to formulate, during a think aloud protocol, an explanation about the puzzling phenomenon after receiving problem-constraining information. Two variables were coded from the transcripts for Phase 2 by two coders blind to the research question and hypothesis; and one variable, final explanation, was coded from the participants' written final answer. First the number of categorical errors made by the participants was coded; inter-rater reliability between two coders for this variable, using 20% of the data, was good ($r = .97$). The second variable of interest was the potential points received by the participants’ explanations; inter-rater reliability between two coders for this variable, using 20% of the data, was good ($r = .98$). Finally, the final answer was coded from the participants’ written explanation by two coders blind to the research questions and hypotheses. The final answer score was computed by averaging both coders’ scores; inter-rater reliability between two coders for this variable was good ($r = 1$).

**Number of errors.** The fourth research question was whether experimentally inducing a convergent thinking mode of thinking would lead to an advantage in making fewer categorical errors during Phase 2. The dependent variable in this analysis, number of errors produced, was tabulated from the coded think aloud data. An ANOVA was run with one between-subjects factor. The between-subjects factor, Condition had three levels: divergent prime (AUT), convergent prime (RAT), and no prime (CON). A main effect of Condition was found, $F(2, 127) = 3.07, p = .05, \eta^2_p = .05$. Given the medium effect size, we opted to choose the LSD post hoc. Post hoc tests using the LSD procedure revealed that participants in the convergent thinking group ($M = 2.75, SD = 2.87$), as shown in Figure
21, exhibited fewer categorical errors on average than did participants in the divergent thinking group ($M = 4.64, SD = 4.50, p = .03$), as well as the control group ($M = 4.57, SD = 4.54, p = .04$), whereas participants in the latter two groups (Control and Divergent thinking condition) did not differ ($p = .94$). Our hypothesis for this fourth research question was corroborated: participants in the convergent thinking group made fewer categorical errors than the two other groups.

![Figure 21. Number of categorical errors exhibited during Phase 2 per condition. The error bars represent 95% CIs.](image)

In addition, to understand the effect of time and condition during the BBRT a 3x3 ANOVA was run with one between-subjects factor and one within-subjects factor (see descriptive statistics Table 7). The between-subjects factor Condition had three levels: divergent prime (AUT), convergent prime (RAT), and no prime (CON). The within-
subjects factor Time had 3 levels formed by dividing the 300 seconds into 3 successive intervals: 0-100 seconds (“100”), 100-200 seconds (“200”), and 200-300 seconds (“300”), within which the total number of categorical errors exhibited during each time interval was tabulated. A repeated measures ANOVA did not find any interaction of time and condition \( (F(4, 254) = .38, p = .82, \eta^2_p = .01) \). The main effect found previously was reproduced (See previous analysis). Finally, there was also a main effect of Time, \( F(2, 254) = 6.04, p < .01, \eta^2_p = .05 \). As Figure 22 shows, participants generated fewer categorical errors with time. This observation was confirmed by the significant linear contrast \( (F(1, 127) = 10.97, p < .001, \eta^2_p = .08) \), which indicates that the number of categorical errors decreased over time intervals. In addition, no significant quadratic contrast was found \( (F(1, 127) = .53, p = .47, \eta^2_p = .01) \).

Table 7.
Descriptive Statistics for Number of Categorical Errors by Time Intervals and by Conditions

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Condition</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-100 Seconds</td>
<td>Divergent Thinking</td>
<td>1.86</td>
<td>2.13</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>1.83</td>
<td>2.08</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
<td>1.02</td>
<td>1.29</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1.57</td>
<td>1.90</td>
<td>130</td>
</tr>
<tr>
<td>100- 200 Seconds</td>
<td>Divergent Thinking</td>
<td>1.61</td>
<td>2.00</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>1.60</td>
<td>1.86</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
<td>0.98</td>
<td>1.52</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1.39</td>
<td>1.82</td>
<td>130</td>
</tr>
<tr>
<td>200- 300 Seconds</td>
<td>Divergent Thinking</td>
<td>1.16</td>
<td>1.48</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>1.14</td>
<td>1.60</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
<td>0.75</td>
<td>1.31</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1.02</td>
<td>1.47</td>
<td>130</td>
</tr>
</tbody>
</table>
Figure 22. Number of categorical errors exhibited during Phase 2, averaged over 100-second intervals. The error bars represent 95% CIs.

Potential awarded points for explanations during think aloud. The fifth research question was whether experimentally inducing a convergent mode of thinking would lead to an advantage in the generation of better explanations during Phase 2. The dependent variable in this analysis, potential awarded points for the generated explanations, was tabulated from the coded think aloud data. First, an ANOVA was run with one between-subjects factor (see descriptive statistics Table 8). The between-subjects factor, Condition had three levels: divergent prime (AUT), convergent prime (RAT), and no prime (CON). No significant result was found $F(2, 127) = 2.14, p = .12, \eta^2_p = .03$. Our hypothesis for this fifth research question was not supported: participants in the convergent thinking group did not generate explanations leading to more potential points than did the two other
groups. In addition, to understand the effect of time and condition during the BBRT, a 3x3 ANOVA was run with one between-subjects factor and one within-subjects factor. The between-subjects factor Condition had three levels: divergent prime (AUT), convergent prime (RAT), and no prime (CON). The within-subjects factor Time had 3 levels formed by dividing the 300 seconds into 3 successive intervals: 0-100 seconds (“100”), 100-200 seconds (“200”), and 200-300 seconds (“300”), within which the potential points for generated explanations during each time interval was tabulated. The Mauchly’s sphericity test was significant, indicating that the assumption of sphericity was violated. Therefore, we used the Greenhouse-Geisser correction. There was a main effect of Time, $F(1.82, 231.02) = 34.14, p < .001, \eta^2_p = .21$. As Figure 23 shows, the explanations that participants generated were awarded fewer points with time. This observation was confirmed by the significant linear contrast ($F(1, 127) = 47.02, p < .001, \eta^2_p = .27$), which indicates that the number of potential points generated decreased over time. In addition, the significant quadratic contrast ($F(1, 127) = 15.76 , p < .001, \eta^2_p = .11$), suggests that potential points generated may have been reaching an asymptote at the end of 300 seconds. Finally, there was no interaction of Condition and Time ($F(3.64, 231.02) = .68, p = .60, \eta^2_p = .01$).
Table 8  
Descriptive Statistics for Number of Awarded Points by Time Intervals and by Conditions

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Condition</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-100 Seconds</td>
<td>Divergent Thinking</td>
<td>0.86</td>
<td>0.93</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>0.98</td>
<td>0.87</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
<td>1.18</td>
<td>1.13</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1.01</td>
<td>0.98</td>
<td>130</td>
</tr>
<tr>
<td>100- 200 Seconds</td>
<td>Divergent Thinking</td>
<td>0.36</td>
<td>0.65</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>0.19</td>
<td>0.51</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
<td>0.34</td>
<td>0.78</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.30</td>
<td>0.65</td>
<td>130</td>
</tr>
<tr>
<td>200- 300 Seconds</td>
<td>Divergent Thinking</td>
<td>0.30</td>
<td>0.70</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>0.12</td>
<td>0.40</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Convergent Thinking</td>
<td>0.30</td>
<td>0.88</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.24</td>
<td>0.69</td>
<td>130</td>
</tr>
</tbody>
</table>

Figure 23. Points awarded for generated explanations during Phase 2, averaged over 100-second intervals. The error bars represent 95% CIs.

**Final score.** The sixth research question was whether experimentally inducing a convergent mode of thinking would lead to an advantage in finding the correct solution,
leading to highest final score, during Phase 2 of the designed scientific reasoning task. The dependent variable in this analysis was the number of points received at the end of Phase 2. An ANOVA was run with one between-subjects factor (see descriptive statistics Table 9). The between-subjects factor, Condition had three levels: divergent prime (AUT), convergent prime (RAT), and no prime (CON). No statistical differences were found between groups ($F(2, 123) = .75, p =.47, \eta^2_p = .01$). Our hypothesis for this sixth research question was not supported: participants in the convergent thinking group did not generate explanations with higher scores than the two other groups.

Table 9

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Divergent Thinking</td>
<td>0.86</td>
<td>0.77</td>
<td>43</td>
</tr>
<tr>
<td>Control</td>
<td>0.83</td>
<td>0.78</td>
<td>40</td>
</tr>
<tr>
<td>Convergent Thinking</td>
<td>0.67</td>
<td>0.68</td>
<td>43</td>
</tr>
<tr>
<td>Total</td>
<td>0.79</td>
<td>0.74</td>
<td>126</td>
</tr>
</tbody>
</table>

Evaluation of the explanation. Our seventh research question suggests that participants in the convergent thinking group should show higher correspondence between the actual accuracy of their explanations and their confidence in their answers in regard to the goal state/solution than the two other groups.

At the end of Phase 2, participants wrote down their final answer as their explanation. These written responses were coded as correct if they identified the two variables of interest (i.e., strength and angle of throw) or incorrect otherwise. Participants also rated their confidence in their explanation on a scale from 0 to 100, where 0 represents
“I don’t think I have the right explanation” and 100 represents “I know I found the correct explanation.” The confidence ratings were dichotomized to form a low-confidence group (rating of 50 or less) and a high-confidence group (rating greater than 50). On this basis, participants were classified into one of four groups (Table 10). Participants’ evaluations were classified as correct evaluation in two cases 1—“Hit”, they answered that they were confident (>50%) in their answer and their answer was correct (They identified the 2 variables of interest); 2 – “Correct Rejection”, they answered that they were not confident (=<50%) in their answer and their answer was not correct (They did not identify the 2 variables of interest). Participants’ evaluations were classified as incorrect evaluation in two cases 1— “False Positive”, they answered that they were confident (>50%) in their answer and their answer was not correct (They did not identify the 2 variables of interest); 2 – “False Negative”, they answered that they were not confident (=<50%) in their answer and their answer was correct (they identified the 2 variables of interest).

Table 10

<table>
<thead>
<tr>
<th>Categorization of Participants’ Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect Explanation</td>
</tr>
<tr>
<td>Correct Explanation</td>
</tr>
<tr>
<td>Low Confidence =&lt;50</td>
</tr>
<tr>
<td>High Confidence &gt;50</td>
</tr>
</tbody>
</table>

*Correct solutions were coded if the participant found the 2 variables of interest (i.e. strength and angle of throw)

Participants were then collapsed into two groups on the basis of whether their evaluations were correct or incorrect. Correct evaluations are 1- when a participant found the two variables of interest (correct explanation) and exhibited high confidence in her explanation (i.e., hit), 2- when a participant did not find the two variables of interest.
(incorrect explanation) and exhibited low confidence in her explanation (i.e., correct rejection). Incorrect evaluations are 1- when a participant found the two variables of interest (correct explanation) and exhibited low confidence in her explanation (i.e., false negative), 2- when a participant did not find the two variables of interest (incorrect explanation) and exhibited high confidence in her explanation (i.e., false positive). Table 11 presents the percentage of participants in each of the three conditions coming to correct vs. incorrect evaluations.

Table 11
Percentage of Participants by Conditions coming to Correct vs. Incorrect Evaluations of their Explanations

<table>
<thead>
<tr>
<th>Condition</th>
<th>Incorrect Evaluation</th>
<th>Correct Evaluation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Divergent Thinking</td>
<td>26 (60.5%)</td>
<td>17 (39.5%)</td>
<td>43 (100%)</td>
</tr>
<tr>
<td>Control</td>
<td>24 (60%)</td>
<td>16 (40%)</td>
<td>40 (100%)</td>
</tr>
<tr>
<td>Convergent Thinking</td>
<td>17 (39.5%)</td>
<td>26 (60.5%)</td>
<td>43 (100%)</td>
</tr>
</tbody>
</table>

Although participants in the convergent thinking group showed a numerical tendency of more often coming to a correct evaluation that did the other two groups, this pattern was not statistically significant ($\chi^2 = 6.05, p = .20$). Our hypothesis suggesting that participants in the convergent thinking group should better evaluate their explanation in regard to the goal state/solution than the two other groups was not supported.

Predicting Scientific Reasoning: Role of Contextual Focus, Intellect and Openness

This section of the analysis investigates the roles played by Contextual Focus (capacity to shift from exploration-divergent thinking- to exploitation-convergent thinking)
and the personality traits of Intellect and Openness during scientific reasoning. In this dissertation, we have collected three different measures of scientific reasoning: 1- the ability to explore the problem space during the BBRT (Phase 1), 2- the ability to exploit the evaluative space during the BBRT (Phase 2), 3- the general ability to reason scientifically using the LTSR.

Our eighth hypothesis was that participants’ performance during exploration of the problem space (Phase 1 of BBRT) should be significantly predicted by their Contextual Focus and Openness scores. The ninth hypothesis was that participants’ performance during exploitation of the evaluative space should be significantly predicted by their Contextual Focus and Intellect scores. Finally, the tenth research question was that participants’ broader score of scientific reasoning (LTSR) should be significantly predicted by their Contextual Focus, Intellect and Openness scores.

Concerning the operational definitions, Intellect and Lawson scores were automatically computed after the participants responded to a computerized Qualtrics version of both measures. Contextual Focus scores were computed by aggregating the two Z-scores obtained for the perceptual-motor convergent thinking task (i.e., Elaborative Reproduction Task [ERT]) and the perceptual-motor divergent thinking task (i.e., Ruff Figural Fluency Task [RFFT]). The motivation to aggregate the Z-scores obtained during the convergent and divergent thinking tasks is to define participants' capacity to appropriately vary their Contextual Focus depending on the task requirements and the predominant task goal (convergent vs. divergent). The proxy is used because a participant who is able to perform highly in both modes of thinking (Convergent and Divergent)
should be more efficient moving between them than another participant who is not very good at either one. Participants performed two different convergent thinking perceptual tasks, the ERT tasks. Each ERT task was fully coded by two coders. Inter-rater reliability between two coders for this variable, using 20% of the data, was good for ERT form A \( r = .95 \) and for form B \( r = .97 \). Participants performed two different divergent thinking perceptual tasks, the RFFT tasks. The same coding process was applied for the two forms of RFFT; inter-rater reliability between two coders for this variable, using 20% of the data, was good \( r = .98 \) for both forms.

**Predicting exploration of problem space performance.** Our eighth hypothesis was that participants’ performance during exploration of the problem space (Phase 1 of BBRT) should be significantly predicted by their Contextual Focus and Openness scores. In order to obtain a general Phase 1 Performance/Ability we computed the score for overall phase 1 performance by aggregating Z-scores of individual participants’ scores obtained during Phase 1 of the BBRT (number of hypotheses, number of categories and switches between categories). The variable Openness was not included because its initial correlation with Phase 1 performance was close to zero (Pearson \( r = -.02 \)) and did not reach significance; in contrast both Intellect \( r = .27, p < .01 \), and the Lawson Score \( r = .33, p < .001 \) were correlated with Phase 1 performance.

A multiple regression model was fitted to the Phase 1 performance variable, where Intellect and Contextual Focus were the predictors. Contextual Focus was added first into the model, followed by Lawson and then Intellect. The results indicated that Contextual
Focus was the strongest predictor of Phase 1 performance \((B = 0.27, t = 2.87, p = .005)\), and that Lawson was also a predictor of Phase 1 performance \((B = 0.04, t = 2.23, p = .03)\). Intellect \((B = 0.03, t = 1.80, p = .07)\) was not a statistically significant predictor of Phase 1 performance (Table 12). The final model explained 19% of the variance in Phase 1 performance. Our hypothesis for this eighth research question was partially corroborated with Contextual Focus being a significant predictor of the problem space exploration ability.

### Table 12

**Regression Model Predicting Phase 1 Performance**

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>(B) (unstandardized)</th>
<th>SE(B)</th>
<th>t-statistic</th>
<th>p-value</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Contextual Focus</td>
<td>0.27</td>
<td>0.095</td>
<td>2.87</td>
<td>.005</td>
<td>19.0%</td>
</tr>
<tr>
<td></td>
<td>Lawson Score</td>
<td>0.04</td>
<td>0.019</td>
<td>2.23</td>
<td>.027</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intellect</td>
<td>0.03</td>
<td>0.014</td>
<td>1.80</td>
<td>.074</td>
<td></td>
</tr>
</tbody>
</table>

Note. Total model \(R^2 = 19.0\%\).

**Predicting exploitation of evaluative space performance.** The ninth hypothesis was that participants’ performance during exploitation of the evaluative space should be significantly predicted by their Contextual Focus and Intellect scores. The additional variable to be computed for this analysis was Phase 2 overall performance. Because the correctness of the explanations is critical during the phase 2 of the BBRT, the Z-scores of the awarded points and the final answer explanations were aggregated to obtain the overall measure of Phase 2 performance.

A multiple regression model using a hierarchical approach was fitted to the Phase 2 performance variable. In Step 1, Phase 1 performance scores were entered as a predictor. In Step 2, Intellect, Contextual Focus, and Lawson scores were added to the model. In both Steps, Phase 1 performance was a significant predictor of Phase 2 performance \((B = 0.25, t = 2.87, p = .005)\).
Neither, Intellect \((B = 0.02, t = 1.23, p = .22)\), nor Contextual Focus \((B = 0.01, t = 0.07, p = .94)\), nor Lawson scores \((B = 0.01, t = 0.35, p = .729)\) were significant predictors of Phase 2 overall performance. The final model explained 11% of the variance in Phase 2 performance (Table 13). Our hypothesis for this ninth research question was not supported.

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>(B) (unstandardized)</th>
<th>SE(B)</th>
<th>t-statistic</th>
<th>p-value</th>
<th>(R^2) Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phase 1 Performance</td>
<td>0.30</td>
<td>0.085</td>
<td>3.56</td>
<td>.001</td>
<td>9.3%</td>
</tr>
<tr>
<td>2</td>
<td>Phase 1 Performance</td>
<td>0.25</td>
<td>0.100</td>
<td>2.59</td>
<td>.011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contextual Focus</td>
<td>0.01</td>
<td>0.108</td>
<td>0.07</td>
<td>.941</td>
<td>1.5%</td>
</tr>
<tr>
<td></td>
<td>Lawson Score</td>
<td>0.01</td>
<td>0.020</td>
<td>0.35</td>
<td>.729</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intellect</td>
<td>0.02</td>
<td>0.015</td>
<td>1.22</td>
<td>.222</td>
<td></td>
</tr>
</tbody>
</table>

Note. Total model \(R^2 = 10.8\%\).

**Predicting broader scientific reasoning performance.** The tenth research question was that participants’ broader score of scientific reasoning (LTSR) should be significantly predicted by their Contextual Focus, Intellect and Openness scores.

The LTSR is a more general test of scientific reasoning consisting in answering multiple choice questions about different aspects of scientific reasoning (i.e. conservation of mass and volume, proportional reasoning, control of variables, probabilistic thinking, correlational thinking, and hypothetical-deductive reasoning). A multiple regression model using a hierarchical approach was fitted to the LTSR variable, which is a measure of general scientific reasoning. In Step 1, Phase 1 and Phase 2 performance scores were entered as predictors. In Step 2, Intellect was added to the model, and finally in Step 3,
Contextual Focus was added to the model. The variable Openness was not included because its initial correlation with LTSR ability did not reach significance.

The Step 1 results indicated that Phase 1 performance was a significant predictor of Lawson scores ($B = 1.56, t = 3.74, p < .001$), while Phase 2 performance was not a significant predictor ($B = 0.29, t = 0.68, p = .50$). Together the two variables accounted for 12.6% of the variance in Lawson scores. The addition of Intellect in Phase 2 accounted for an additional 4.5% of the variance in Lawson scores; the Intellect variable was statistically significant individually ($B = 0.18, t = 2.55, p = .012$). After the addition of the Intellect variables, the Phase 1 performance variable was still a statistically significant predictor of Lawson scores ($B = 1.30, t = 3.04, p = .003$).

Finally, the addition of Contextual Focus to the model in Step 3 accounted for another 4% of the variance in Lawson scores (final model $R^2 = .21$). In the final model, Phase 1 performance was still a statistically significant predictor of Lawson scores ($B = 0.90, t = 2.04, p = .043$). Both Intellect ($B = 0.17, t = 2.50, p = .014$) and Contextual Focus ($B = 1.15, t = 2.45, p = .016$) were significant predictors of Lawson scores in the final model (Table 14). Our hypothesis for this tenth research question was partially corroborated, with the exception of openness not being a good predictor of general scientific reasoning ability.
Table 14
*Regression Model Predicting LTSR*

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>B(unstandardized)</th>
<th>SE(B)</th>
<th>t-statistic</th>
<th>p-value</th>
<th>$R^2$ Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phase 1 Performance</td>
<td>1.56</td>
<td>0.417</td>
<td>3.74</td>
<td>&lt;.001</td>
<td>12.6%</td>
</tr>
<tr>
<td></td>
<td>Phase 2 Performance</td>
<td>0.29</td>
<td>0.423</td>
<td>0.68</td>
<td>.500</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Phase 1 Performance</td>
<td>1.28</td>
<td>0.422</td>
<td>3.04</td>
<td>.003</td>
<td>4.5%</td>
</tr>
<tr>
<td></td>
<td>Phase 2 Performance</td>
<td>0.16</td>
<td>0.416</td>
<td>0.37</td>
<td>.709</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intellect</td>
<td>0.18</td>
<td>0.069</td>
<td>2.56</td>
<td>.012</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Phase 1 Performance</td>
<td>0.90</td>
<td>0.442</td>
<td>2.04</td>
<td>.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phase 2 Performance</td>
<td>0.14</td>
<td>0.408</td>
<td>0.35</td>
<td>.729</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intellect</td>
<td>0.17</td>
<td>0.067</td>
<td>2.50</td>
<td>.014</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contextual Focus</td>
<td>1.16</td>
<td>0.472</td>
<td>2.45</td>
<td>.016</td>
<td></td>
</tr>
</tbody>
</table>

Note. Total model $R^2 = 21.0\%$. 

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>B(unstandardized)</th>
<th>SE(B)</th>
<th>t-statistic</th>
<th>p-value</th>
<th>$R^2$ Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phase 1 Performance</td>
<td>1.56</td>
<td>0.417</td>
<td>3.74</td>
<td>&lt;.001</td>
<td>12.6%</td>
</tr>
<tr>
<td></td>
<td>Phase 2 Performance</td>
<td>0.29</td>
<td>0.423</td>
<td>0.68</td>
<td>.500</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Phase 1 Performance</td>
<td>1.28</td>
<td>0.422</td>
<td>3.04</td>
<td>.003</td>
<td>4.5%</td>
</tr>
<tr>
<td></td>
<td>Phase 2 Performance</td>
<td>0.16</td>
<td>0.416</td>
<td>0.37</td>
<td>.709</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intellect</td>
<td>0.18</td>
<td>0.069</td>
<td>2.56</td>
<td>.012</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Phase 1 Performance</td>
<td>0.90</td>
<td>0.442</td>
<td>2.04</td>
<td>.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phase 2 Performance</td>
<td>0.14</td>
<td>0.408</td>
<td>0.35</td>
<td>.729</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intellect</td>
<td>0.17</td>
<td>0.067</td>
<td>2.50</td>
<td>.014</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contextual Focus</td>
<td>1.16</td>
<td>0.472</td>
<td>2.45</td>
<td>.016</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 7: Discussion

The broader goal of this dissertation was to bridge the process of scientific reasoning with the field of cognitive science, more particularly, the cognitive mechanisms involved during reasoning. Although the traditional view of scientific reasoning incorporates different steps or facets in the reasoning process (i.e. the “Alpha” model), the scientific reasoning literature lacks a more cognitively-grounded approach. The cognitive dualism of exploration/exploitation, that is key to understanding the steps involved in seeking scientific truth, is not represented or explicitly considered by our current traditional representation of scientific reasoning. The proposed “Z” model of scientific reasoning embodies the exploration of the problem space as a divergent thinking enterprise and the exploitation of the evaluative space as a convergent thinking enterprise.

Convergent and divergent thinking have been identified as playing a critical role in complex problem-solving tasks (DeYoung, Flanders & Peterson, 2008; Guilford, 1956; Guilford, 1959; Guilford, 1967; Hommel, 2012). This intent of bridging scientific reasoning with cognitive mechanisms resulted in a new model: the "Z" model of scientific reasoning. The "Z" model integrates the traditional scientific reasoning steps while depicting the cognitive mechanisms and mental flexibilities at use during reasoning. In addition, the mental shifting between convergent and divergent thinking, here termed Contextual Focus, is represented in the “Z” model as the central mental flexibility allowing the individual to shift between exploration and exploitation during reasoning.
scientific reasoning. Individual differences such as personality traits (i.e. Intellect and Openness) related to convergent and divergent thinking were also accounted for when testing the role of Contextual Focus during scientific reasoning.

The specific goals of this research were to empirically test the “Z” Model with three broad sets of predictions:

1. If the exploration of the problem space is a divergent thinking enterprise, then participants experimentally primed to think divergently during this phase should display an advantage compared to the participants not so primed,

2. If the exploitation of the evaluative space is a convergent thinking enterprise, then participants experimentally primed to think convergently during this phase should display an advantage compared with participants not so primed,

3. If Contextual Focus is a key ability in navigating between mental exploitation and exploration, and if Openness and Intellect are personality traits related to divergent and convergent thinking, then these three variables should play a role during scientific reasoning. In other words, a measure of Contextual Focus performance, Intellect and Openness scores should be good predictors of scientific reasoning performance.

As we just mentioned, three distinct clusters of research questions were identified for this dissertation, therefore the findings and their associated research questions will be presented in three sub-sections: 1- Exploration of the problem space, 2- Exploitation of the evaluative space, 3- Predictive power of Contextual Focus, Intellect, and Openness during scientific reasoning. Finally, we will close this section with a discussion about additional broader cognitive considerations in light of our findings.
Exploration of the Problem Space. Regarding the exploration of the problem space (P1 of the BBRT), our larger research question was: “Does any type of cognitive experimental priming lead to an advantage when exploring the problem space?” Our first hypothesis was corroborated by the results; participants in the divergent thinking group generated on average more hypotheses ($M = 14.58$) than did the participants in the convergent thinking group ($M = 10.41$) and the control group ($M = 11.12$). In addition, the observed linear relationship between the number of hypotheses generated and time interval during P1 posits that, regardless of their experimental condition: as the amount of time spent during phase 1 increased, the number of hypotheses generated decreased. In addition, the quadratic relationship between the number of hypotheses generated and time suggests that participants had enough time during our experimental protocol to express their explanations about the puzzling phenomenon. This finding also resonates with a previous study (Chrysikou, 2006) where categorization (e.g. divergent thinking effort) and goal directed cognition (e.g. convergent thinking effort) were described as the dynamic expression of problem solving. Our finding, highlighting the cognitive advantage of the divergent thinking group, is aligned with Chrysikou’s finding about the importance of categorization. We found that the divergent thinking group outperformed the two other groups in generating variables.

Our second hypothesis—that participants in the divergent thinking group should generate more categories of variables than the two other groups—was not supported. The reason for this non-finding could be related to the small number of potential categories to be generated (i.e., four; Action on the Ball, Environment, Human Factor, Ball difference).
A lack of variability arising from a ceiling effect is a possible explanation given that, on average, participants were able to generate 3.36 categories out of the four.

Finally, regarding the exploration of the problem space (BBRT - Phase1), our third hypothesis—that participants in the divergent thinking group should switch between categories of variables more often than the two other groups—was also not supported. The previous arguments elaborated regarding the lack of significance when generating categories, are here also relevant.

In sum, “Does any type of cognitive experimental priming lead to an advantage when exploring the problem space?” Our first hypothesis’ finding suggests that—as depicted in the “Z” model—divergent thinking mode is the mental mechanism of choice during problem space exploration (Figure 24). Participants exhibited more extensive generation of hypotheses after being experimentally divergently primed. This finding partially validates the “Z” Model and extends Klahr and Dunbar’s (1988) work on scientific reasoning mental space searches. In addition to knowing that reasoning styles (individual differences) exist regarding scientific reasoning—The theorists favor revisiting the hypothesis space before finding the correct solution (Klahr & Dunbar, 1988) – we can posit now that experimental manipulation, divergent thinking priming, can enhance scientific reasoning: in particular, the exploration of the problem space. Another scientific reasoning study also found, using different measures with a younger population than was tested by Klahr and Dunbar (1988), that children who happened to be the best at scientific reasoning tasks, were the ones that were able to evaluate their evidence (e.g. “experimenters”) and at the “same time,” to evaluate their changing
theories (e.g. “theorists”) (Schauble, 1990). Whereas Klahr and Dunbar (1988), as well as Schauble’s work (1990) on scientific reasoning, seems to describe scientific reasoning inquiry as a trait-based attribute (e.g. “experimenters” vs. “theorists”), our findings underscore the contribution of a state-based factor to scientific reasoning. That is, although an individual may have a general (longer-term) trait predisposition toward either experimenting or theorizing, on a briefer timescale, specific contextual factors – such as the nature of the prior tasks one has been engaged upon – can also momentarily enhance participants’ performance during specific phases of scientific reasoning.

Some within-group additional findings also suggest that the exploration of the problem space is related to divergent thinking performance during the priming phase: participants who generated more ideas during the AUT on average also displayed higher Phase 1 performance ($r = .56, p < .001$). Concerning this first research question about scientific reasoning, we found that divergent cognitive experimental priming lead to an advantage – more hypotheses generated – when exploring the problem space. In addition, in the control group, participants who generated on average more words also displayed a higher Phase 1 performance ($r = .30, p < .05$). In contrast, in the convergent thinking group, the RAT average score was not correlated with Phase 1 performance ($r = .08, p = .61$). This differential pattern of correlations does not only confirm the divergent thinking nature of the task during the exploration of the problem space, it also validates the choice for the experimental priming tasks.
Figure 24. Problem Space Exploration in the “Z” Model of Scientific Reasoning

Exploitation of the Evaluative Space. Regarding the exploitation of the problem space (BBRT - Phase2), our larger research question was: “Does any type of cognitive experimental priming lead to an advantage when exploiting the problem space?” Our fourth hypothesis was corroborated; participants in the convergent thinking group displayed on average fewer categorical errors ($M = 2.75$) than the participants in either the divergent thinking group ($M = 4.64$) or the control group ($M = 4.57$). In addition, the linear relationship between the number of categorical errors generated and time interval during Phase 2 suggests that as the amount of time spent during Phase 2 increased, participants displayed fewer categorical errors. The study we mentioned earlier (Chrysikou, 2006) suggesting that problem solving is the dynamic expression of categorization (e.g. divergent thinking effort) and goal-directed cognition (e.g. convergent thinking effort), is again relevant to the effect of condition that we observed. The convergent thinking group outperformed the two other groups in goal-directed cognition, by focusing on the correct category of variables given the newly constraining information they had been provided.
Our fifth hypothesis—that participants in the convergent thinking group should generate explanations leading to more potential awarded points than the two other groups—was not supported. The direction of the result, such that participants in the convergent thinking group were on average awarded more points than the two other groups, coupled with the medium effect size played by condition over the awarded points, suggest that a statistically significant difference could be found with a larger sample size. A statistical power limitation could explain the non-significant result. In addition, we also need to consider the eventuality that the “Z” model, and more particularly the exploitation of the evaluative space, may only be reflected by an enhanced ability to avoid generating categorical errors and not extend to an enhanced ability to generate correct explanations.

Our sixth hypothesis—that participants in the convergent thinking group should provide better final answers/explanations than the two other groups—was not supported. Consider that the awarded points during Phase 2 were collected via think-aloud protocol, while the final explanation was written down. The convergent thinking mode related to producing a written answer may have deflated the intended experimental priming and caused the control and divergent thinking group to shift into a more convergent thinking mode, resulting in no significant difference in the accuracy of the final explanation.

Our seventh hypothesis—that participants in the convergent thinking group should better evaluate their explanation in regards to the goal state/solution in comparison to the other two groups—was not supported. A lack of statistical power may again be a factor in this non-finding. The direction of the data suggests a clear advantage for the
convergent thinking group in being able to correctly evaluate their explanation (60%) versus the control group (40%) and the divergent thinking group (39.5%). We believe that a larger sample size could lead to statistically significant difference and corroborate our hypothesis.

In sum, “Does any type of cognitive experimental priming lead to an advantage when exploiting the evaluative space?” Our finding, corroborating our fourth hypothesis, suggests that – as depicted in the “Z” model – convergent thinking mode is the mental mechanism of choice during evaluative space exploitation (Figure 25). Participants exhibited fewer categorical errors after being experimentally convergently primed. This finding partially validates the “Z” Model and extends Klahr and Dunbar’s work on scientific reasoning mental space searches. In addition to knowing that reasoning styles (individual differences) exist regarding scientific reasoning – The experimenters reached the right solution by emphasizing their search in the experimental space (Klahr & Dunbar, 1988) – we can posit now that experimental manipulation, convergent thinking priming, can enhance scientific reasoning: more specifically, the exploitation of the evaluative space. Concerning the mistakes related to valid scientific reasoning, we earlier contemplate the type “d” errors (i.e. “Alpha” model) – due to an error made during the evaluation phase – and our finding reveals that convergent thinking priming helps reducing the occurrence of this particular type of mistake.

Some within-group additional findings also suggest that the exploitation of the evaluative space is related to convergent thinking performance. Specifically, participants who generated more correct answers on the RAT on average also displayed higher Phase
2 performance \((r = .38, p = .01)\). In addition, in the control group, participants who generated more words also displayed a higher Phase 2 performance \((r = .44, p < .01)\). In contrast, in the divergent thinking group, the average AUT score was not correlated with Phase 2 performance \((r = -.19, p = .23)\). This pattern of differential correlations does not only confirm the convergent thinking nature of the task during the exploitation of the evaluative space, it also validates the choice for the experimental priming tasks.

It is also important to note here that our findings related to the exploration of the problem space (i.e. theorists) and the exploitation of the evaluative space (i.e. experimenters) are aligned to some of Schauble’s work on the development of scientific reasoning. Schauble’s description of the successful child (1991): “The most successful children evaluated both the evidence and their changing theories, and were sensitive to the fact that they should be mutually constraining.” This description closely resonates with our findings and the structure of the “Z” model. There are two mental spaces to visit and navigating both, in a dynamically interleaved manner – such that they become "mutually constraining" – is an advantage.

*Figure 25. Evaluative Space Exploitation in the “Z” model of Scientific Reasoning*
Scientific Reasoning: Role of Contextual Focus, Intellect and Openness. Our larger research question for this cluster was: “Can individual Contextual Focus, Intellect and Openness scores predict performance on scientific reasoning tasks?” To fully answer this research question we considered the specific problem space exploration (Phase 1 of the BBRT) and the evaluative space exploitation of scientific reasoning (Phase 2 of the BBRT), as well as a more general performance measure of scientific reasoning (the LTSR). Regarding the predictive power of Contextual Focus and Openness scores during the exploration of the problem space, our eighth hypothesis – scientific reasoning ability during exploration of the problem space should be significantly predicted by Contextual Focus and Openness – was partially corroborated. While controlling for participants’ broader performance to reason scientifically (as assessed by the LTSR), Contextual Focus was found to be a significant predictor of the overall performance in exploring the problem space of our specific scientific reasoning problem (Phase 1 of the BBRT). The very specific nature of the BBRT—consisting of a Newtonian physics-mechanical problem—may have negatively influenced the effect of personality traits such as Openness given that individuals high in Openness typically are more interested in the creative process by seeking original ideas. Openness was not included in any of the predictive models due to its lack of statistical correlation with any of the scientific reasoning tasks (P1 and P2 of the BBRT, and the LTSR). The nature of the scientific task, in comparison to a creative or aesthetically-engaging task, for example, may be less favorable to detect any predictive power of Openness. This non-finding is related to Kaufman and colleagues (2016) stipulating that Openness was more related to creative
achievement in arts, whereas Intellect was more related to creative achievement in science.

Our ninth hypothesis suggesting that participants’ performance during exploitation of the evaluative space should be significantly predicted by their Contextual Focus and Intellect scores was not supported. The nature of our scientific reasoning task (BBRT) may be too specific and the performance during its Phase 2 (P2 of the BBRT) may be too dependent on the performance during Phase 1 (P1 of the BBRT), which in turn, may have limited our capacity to find significant predictive power of individual differences such as Contextual Focus and Intellect.

Our tenth hypothesis – Broader test-based scores of scientific reasoning should be significantly predicted by Contextual Focus, Intellect and Openness – was partially corroborated. After controlling for participants’ performance during our more specific scientific reasoning task (Phase 1 and Phase 2 of the BBRT), Contextual Focus and Intellect were found to be significant predictors of the broader scientific reasoning (LTSR) performance.

In sum; “Can individual Contextual Focus, Intellect and Openness scores predict performance on scientific reasoning tasks?” Concerning the overall third research question about the roles of Contextual Focus, Intellect and Openness during scientific reasoning: except for the prediction of our specific scientific reasoning task performance during Phase 2 (i.e. ninth hypothesis), our findings suggest that – as depicted in the “Z” model – Contextual Focus is related to scientific reasoning performance (Figure 26). Contextual Focus, the shift between exploration of the problem space and exploitation of
the evaluative space is related to the performance during the exploration of the problem space (P1 of BBRT), and is also a significant predictor to overall performance during general, broader scientific reasoning task (LTSR). We also found that Intellect, as hypothesized, is also a significant predictor of general scientific reasoning performance. These findings only partially validate the “Z” Model. While our findings support the role of Contextual Focus during the exploration of the problem space (P1 of BBRT) and the overall broader scientific performance (LTSR), they do not support the role of Contextual Focus during the exploitation of the evaluative space (P2 of BBRT). It is not clear if Contextual Focus plays the central role as presented in the “Z” model. In the realm of scientific reasoning, Contextual Focus could be considered more holistically and not as a clear switch between convergent thinking (i.e. executive central network) and divergent thinking (i.e. default network). Thus, we may want to consider this type of cognitive flexibility (i.e. Contextual Focus) more holistically, as suggested by research on creativity: during creative thinking the executive central network and the default mode network have been observed to be concurrently recruited, or activated at the same time (Beaty, Benedek, Silvia & Schacter, 2016). We may hypothesize that while engaged in creative thinking or scientific reasoning, higher Contextual Focus: 1- allows individuals to search more deeply both spaces simultaneously (i.e. problem and evaluative space), and 2- gives individuals more mental search power to explore the problem space (i.e. divergent thinking – default network activation – resulting in higher Originality), while still being capable of modulating the top-down focus (i.e. convergent thinking –
executive central network) to avoid selecting useless or non-promising ideas (i.e. maintaining a high level of appropriateness).

![Figure 26. The “Z” model of scientific reasoning](image)

**Additional Cognitive Considerations.** Again, while the proposed “Z” model was partially supported by our findings – the generation of hypotheses during the exploration of the problem space, the generation of fewer categorical errors during the exploitation of the problem space, and the critical role of Contextual Focus during scientific reasoning – those findings can also be interpreted with broader cognitive science lenses. What insights can we draw from our experimental manipulations over broader cognitive mechanisms during problem solving?

The bottom-up *perception* consisting in helping us formulating hypotheses can be enhanced by experimentally priming individuals into a more divergent way of thinking.
Regarding the top-down action consisting in focusing on the correct category of variables after receiving feedback from experiments can also be enhanced, by experimentally priming individuals into a more convergent way of thinking. Given that complex mental tasks such as problem solving equally require both divergent and convergent thinking, it is interesting to hypothesize that the priming used during our experimental protocol should also lead to an advantage during more general problem solving. Although Contextual Focus appears to be a predictor of the exploration of the problem space and more general scientific reasoning performance, our findings do not disclose any clear Contextual Focus advantage over the exploitation of the evaluative space. Contextual Focus could be therefore considered as a broader individual difference component acting upon our capacity to reason scientifically. In the light of the broader problem solving arena, Contextual Focus seems to be a good contender to explain individual differences, more especially if the type of problem solving requires a deeper exploration of the problem space and overall scientific thinking.

Category generation and shifting between categories during exploration of the problem space is related to the spontaneous flexibility of the divergent thinking search (Guilford, 1959); However, the fact that only the generation of ideas was found to be significantly different, but not the categories, nor the number of switches, after our divergent thinking priming intervention may also suggest that 1- the divergent thinking priming protocol had a stronger effect on the Fluency rather than the Spontaneous Flexibility component of divergent thinking, or 2- the exploration of the problem space
and more specifically the generation of hypotheses is more sensitive to fluency than Spontaneous Flexibility.

Findings related to the exploitation of the evaluative space (i.e. convergent thinking priming is only associated with fewer categorical errors, but not with awarded points, nor final score) suggest that maybe the effect of convergent thinking priming is not finding a better, an optimal solution; but only its outcome. When the goal is to find only one or the best solution, the level of focus is at the highest because, unlike with divergent thinking goal, a mistake is possible. The activation of the executive central network during convergent thinking (Beaty, Benedek, Silvia & Schacter, 2016) suggests that executive functions such as inhibition, working memory would be solicited. Therefore, this could explain why the level of reactive flexibility (e.g. avoiding categorical mistakes) would be more sensitive to convergent thinking experimental priming during the exploitation of the evaluative space. While the generation of a correct answer may only be a consequence, an outcome, of the effect of convergent thinking mental priming.

Limitations

The first limitation of this research is the size of our sample. A larger sample size may have unveiled a larger number of statistically significant findings. Several statistical tests —such as the conditions ‘differences in the number of switches during the exploration of the problem space (Phase 1 of the BBRT), or the points awarded during the exploitation of the evaluative space (Phase 2 of the BBRT), or again the difference in
the correct evaluation of explanations after the exploitation of the evaluative space (Phase 2 of the BBRT)—would have benefited from a larger sample size given that the directionality of the results matched expectations and also produced small to medium effect sizes.

The nature of our main scientific reasoning task (the BBRT) can also appear as a challenge in terms of corroborating some of our hypotheses. The small number of potential categories, and the ceiling effect observed in the number of categories generated made differences between the experimentally-assigned groups more difficult to observe. This research would benefit from being replicated using different scientific reasoning tasks other than the BBRT. New scientific reasoning tasks requiring the generation of hypotheses belonging to more categories would be beneficial. It might also be beneficial to incorporate a scientific reasoning task focused on a biological or social science phenomena. Inclusion of such a scientific reasoning task could both test the generalizability of the findings reported here, and test whether the personality trait of Openness to Experience might exert a stronger influence when the scientific subject matter calling for divergent thought is less directly tied to the physical sciences. Another limitation regarding the measure of Openness is that it contains only 10 items. Openness may be a more multileveled trait, it may also differ from Intellect on its measurement sensitivity score. Measuring Intellect with only ten items, may be sufficient to acquire a relevant score; However, the construct of Openness is a multifaceted construct that could not be captured by questions only related to intellect or culture (McCrae & Costa, 1985). The BFAS measure of Openness, with only ten items, may have been too limiting.
Contextual Focus scores were not correlated with any of the scores measured during Phase 2 (i.e. convergent thinking goal). However, Contextual Focus scores were correlated with every single score measured during the phase 1 of the BBRT. Such pattern of findings could suggest that contextual focus play a more important role during the exploration of the problem space? Or maybe our measure of Contextual Focus is more inclined towards measuring divergent thinking skills. In addition, the additive nature of convergent and divergent perceptual performance as a “proxy” for Contextual Focus (the shift between convergent and divergent thinking) is debatable. Some may rightly say that our measure of Contextual Focus cannot measure a shift because the Contextual Focus tasks, we use, do not invite the participants in shifting. The idea of interleaving each task (e.g convergent-divergent-convergent-divergent) during the Contextual Focus measure could invite the participant to shift more.

Unlike other experimental cognitive neuroscience paradigms that use, for example, transcranial magnetic stimulation to temporarily stimulate or block the cognitive-behavioral contributions of particular areas of the cortex, our experimental paradigm, consisting in priming participants into a more convergent or more divergent thinking mode, did not prevent participants from switching from one mode to another. We only “suggested” (through a procedural priming intervention) a specific mode of thinking to the participants, and the priming may have had different impact from one participant to another. The durability of the priming intervention, across intervening tasks and time, is also unknown, though it has been proposed to last not more than 45-50 minutes (Wen, Butler, & Koutstaal, 2013, p. 114).
In terms of the corroborated hypotheses, this study design does not affirm that being primed in one specific mode of thinking would lead to an absolute advantage during scientific reasoning. Our findings posit that divergent thinking priming leads to an advantage during the problem space exploration (successful generation of a higher number of hypotheses) and convergent thinking priming leads to an advantage during the exploitation of the evaluative space (the commission of fewer categorical errors given problem-constraining information). However, we cannot claim that priming students in only one specific mode, or of closely and fully matching the modes of thinking to the temporally-changing demands of the problem, would be beneficial for overall scientific reasoning performance.

**Future Directions**

This section will be divided into two parts. First, we will propose a study that addresses the limited scope of our present findings, leading to a potentially more applied educational outcome. Secondly, a broader conceptual point about the “Z” model itself is treated; by incorporating two mental spaces described in the “Alpha” model but yet not represented in the current proposed “Z” model of scientific reasoning (Hypothesis and Experimental space searches).

**Future Study.** While our primary goal for this research was to test the “Z” model of scientific reasoning, it would be informative in future studies to test a few new hypotheses with more applied educational lenses:
1. Priming participants divergently before the exploration of the problem space and priming them convergently before the exploitation of the evaluative space should be beneficial for their overall scientific reasoning performance.

2. Although the hypothesis regarding the evaluation of the explanation (correct evaluation of their explanation) only approached statistical significance, it would be interesting to further test if this result was driven by convergent thinking priming during the entire task (Phase 1 and Phase 2 of the BBRT) or if the same pattern of results would be observed if the participants were only (that is, exclusively) primed convergently before the exploitation of the evaluative space.

In order to test these hypotheses, we propose a new study design comprised of four experimental conditions. To test the first research question, we would compare participants’ performance on the BBRT (Phase 1 and Phase 2) under two conditions. The first condition will consist in control priming (i.e. WA) the participants before each of the phases of the BBRT. The second condition will consist in priming the participants divergently (i.e. AUT) before the exploration of the problem space (i.e. Phase 1 of the BBRT), and convergently (i.e. RAT) before the exploitation of the evaluative space (i.e. Phase 2 of the BBRT). To test the second research question, we would compare participants’ performance during the evaluation of their explanation under two further conditions. Participants in condition 3 would be primed convergently before each phase of the BBRT, whereas, participants in condition 4 will be primed with the control task before the exploration of the problem space (i.e. Phase 1 of the BBRT), then they will be
primed convergently before the exploitation of the evaluative space (i.e. Phase 2 of the BBRT).

This proposed future study is a direct educational application of the “Z” model. Such a study would have implications for the teaching of scientific reasoning, but also for behaviorally-based ways to encourage more efficient critical and scientific reasoning in general. Such study, in case of positive results, would be an example of how the “Z” model could be applied to understand and enhance scientific reasoning performances. Such a study would be in line, using our current findings, to test a more ecological application of scientific reasoning. Concerning the broader contribution of our findings to the field of scientific reasoning, we hope that the Z-model, as well as the experimental priming, will be replicated to better understand how we can cognitively enhance scientific reasoning ability. A study about the development of scientific reasoning posits that both valid strategies for generating and interpreting evidence (e.g. exploitation of the evaluative space), and correct beliefs (e.g. exploration of problem space) were not, alone, a sign of success (Schauble, 1996). It will be interesting to find out if such model of scientific reasoning development is related to participants’ performances during our paradigm. In “Schauble’s words” (1996), finding the correct solution (e.g. BBRT), will be dependent upon the participants’ capacities to explore the problem space and exploit the evaluative space. Our condition where participants are primed divergently before the exploration phase, then primed convergently before the exploitative phase, should perform better than the other groups.

127
Toward a Fine-Grained Depiction of the “Z” Model. Since the “Z” model tested in this research did not fully represent the ““Alpha”” Model of scientific reasoning, we present in this last section the two intermediate space searches of the ““Alpha”” model, along with their graphical depictions: 1 – the hypothesis space search, marking the end of the exploration phase of the problem space, 2 – the experimental space search, marking the beginning of the exploitation of the evaluative space.

Hypothesis Space Search is part of the larger problem space exploration; nonetheless, the search in hypothesis space search requires more complex reasoning to formulate and select the best hypotheses. As we previously mentioned, complex reasoning requires both convergent and divergent thinking, and the associative nature of the hypothesis search does not depart from the rule. The hypothesis space search consists of using the operators and variables initially identified during the problem space exploration and combining them through mental associations to generate plausible hypotheses. The associative nature of the task and the almost automatic selection of viable associations (hypotheses) reflects the explorative-exploitative nature of the hypothesis space search (Figure 27). Associative thinking and more particularly analogical reasoning have been found to be frequently present in scientific reasoning (Dunbar, 2000). Please note that during this research we asked participants to generate hypotheses, but we did not have an experimental design task requiring them to select the one they would most like to test. As we defined earlier (p.15), abductive reasoning consists of generating hypotheses using the operators and variables identified during the
problem space search, while retroductive reasoning is the mental testing/monitoring of those hypotheses in accordance with our current knowledge (Lawson, 2010).

Figure 27. Hypothesis Space Search: Exploration & Exploitation

**Experimental space search.** After completing a divergent-thinking goal, exploring the problem space and the hypothesis space search, we move to a more convergent thinking goal (e.g. "pull the net back into the boat"). After selecting one or multiple hypotheses to test, we move from the exploring phase to the exploiting phase. The experimental space search (Figure 28) consists of thinking about the research design, the next step after the selection of a valid-appropriate hypothesis. We could translate this into our fishing analogy by stating that after having chosen what type of fish you want (hypothesis space search), you prepared an appropriate net to catch them (experimental space search), then you cast it in the ocean and pull the cord (end of the evaluative space exploitation). The complex experimental space search is also composed of an exploration
and exploitation phase, like the broader model of scientific reasoning, and the hypothesis space search. "What can I do?" and "What are the tools that I have to test this hypothesis in the real world?" are some questions that summarize the exploration part of the experimental space search. Knowledge of experimental design, such as the need to control for variables or including one or more control groups in the design, will be explored during the associational phase of the experimental search. In the same way that we earlier generated hypotheses and discarded others, we are now exploring some of our research design tools that we found appropriate to test our chosen hypothesis, and we exploit their validity-appropriateness given the context of our reality: "How should I do it?" and "What are the constraints that I will face to test this hypothesis." Questions of this form will help us exploit the experiment space search and ultimately finalize our research design when engaging in scientific reasoning.

Figure 28. Experimental Space Search: Exploration & Exploitation
“Z” Model 2.0. The aim of this dissertation was to introduce and test a new model of scientific reasoning: the “Z” model. As we discussed earlier, the model tested in this research is incomplete in regards to the “Alpha” model of scientific reasoning. In this last part, we present a more complete graphical representation of the entire “Z” model. The reason for situating the hypothesis space search within the larger exploration of the problem space and the experimental space search within the larger exploitation of the evaluative space, is due to their respective goals within the scientific reasoning endeavor. The perception-action cycle is key to categorize the types of search that takes place in the mind after perception (bottom-up triggering of the exploration of problem space) and the types of searches that aim to optimize the action (top-down purpose of the exploitation of the evaluative space) to test and evaluate our hypothesis in the outside world. On the one hand, the hypothesis space search belongs to the problem space search because of the internal (perception-mind) divergently motivated type of search it represents. On the other hand, the experimental space search belongs to the evaluative space search because of the external (action- outside world) convergently motivated type of search it serves (Figure 29).
Figure 29. “Z” Model 2.0

The “Alpha” model is the traditional view to depict scientific reasoning. The “Z” model includes the same mental space searches, but in addition it depicts the types of cognitive modes that are dominant for each type of mental search (e.g. exploration of problem space), or sub-mental search (e.g. retroduction). The development of such a model is not only beneficial for the understanding of scientific reasoning, but we argue that it can also be applied to the science classroom to improve student’s scientific reasoning. Finally, because scientific reasoning is a complex cognitive task, we claim that the “Z” model should also be considered as a general cognitive model of reasoning, that can provide insights into the different cognitive mental space searches for a variety of complex cognitive tasks, including other forms of complex or ill-defined problem-solving, critical thinking and reasoning, and creative thinking in its many and diverse manifestations in our individual and collective lives.
References


138


Thinking and CogFLEX, a New measure of Cognitive Flexibility.

Poster presented at: Association for Psychological Science. 27th Annual Convention; 2015 May 21-24; New York, NY.


Appendix A

Scientific reasoning task: The Bouncing Balls Reasoning Task

Phase 1

*Imagine you are a middle school science teacher. One of your students, Pat, saw a red ball going 8 feet high in the air after the first bounce on the floor, and it stopped after 6 bounces.*

*A few minutes later Pat saw a blue ball going 4 feet high in the air after the first bounce on the floor, and it stopped after 12 bounces.*

During the next 5 minutes please tell us why you think the phenomenon described above is possible? In other words, please formulate **as many hypotheses** as you can to explain the described phenomenon.”

Phase 2

After running different experiments you and your students conclude two things:
- **The environment, such as the floor and weather conditions, did not play any role in the phenomenon previously described.**
- **Except for their color, the balls are strictly identical.**

Please formulate an explanation of why one ball went 8 feet high in the air after the first bounce on the floor, and it stopped after 6 bounces whereas the other ball went 4 feet high in the air after the first bounce on the floor and it stopped after 12 bounces.