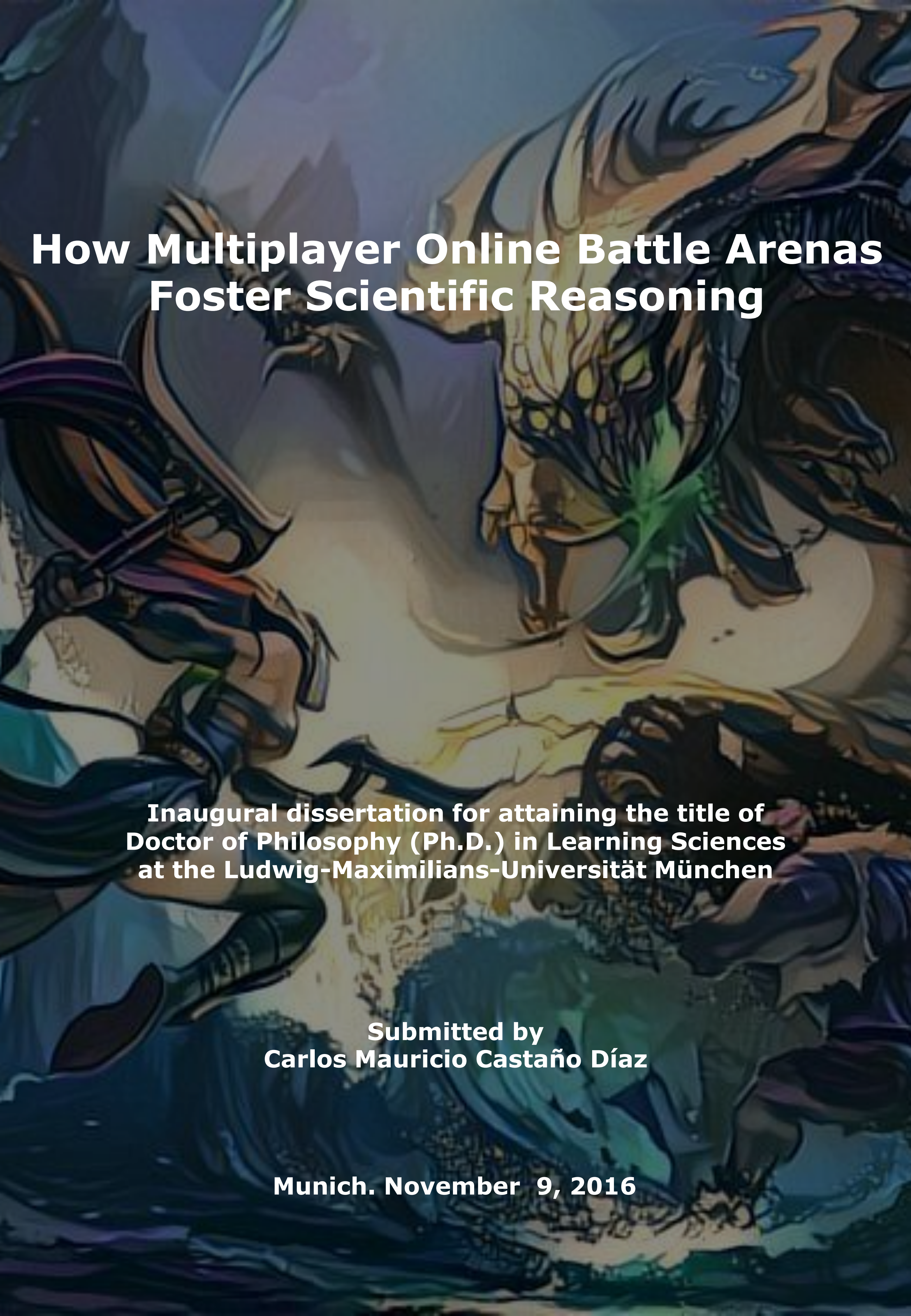


How Multiplayer Online Battle Arenas Foster Scientific Reasoning

Inaugural dissertation for attaining the title of
Doctor of Philosophy (Ph.D.) in Learning Sciences
at the Ludwig-Maximilians-Universität München

Submitted by
Carlos Mauricio Castaño Díaz

Munich. November 9, 2016

How Multiplayer Online Battle Arenas Foster Scientific Reasoning

Dissertation zum Erwerb des Doctor of Philosophy (Ph.D.)
am Munich Center of the Learning Sciences der
Ludwig-Maximilians- Universität München

Submitted by
Carlos Mauricio Castaño Díaz
in Munich

Munich. November 9, 2016

First Supervisor: Prof. Dr. Birgit Dorner
Second Supervisor: Prof. Dr. Heinrich Hussmann
Third Supervisor: Prof. Dr. Jan-Willem Strijbos
International Supervisor: Prof. Dr. Richard Allan Bartle

Date of the oral defence February the 3rd 2017

Eidesstattliche Versicherung

Statement of Scientific Integrity

Name, Vorname

Last name, first name

Ich versichere, dass ich die an der Fakultät für Psychologie und Pädagogik der Ludwig-Maximilians-Universität München zur Dissertation eingereichte Arbeit mit dem Titel:

I assert that the thesis I submitted to the Faculty of psychology and pedagogy of the Ludwig-Maximilian-Iniversitaet Muenchen under the title:

selbst verfasst, alle Teile eigenständig formuliert und keine fremden Textteile übernommen habe, die nicht als solche gekennzeichnet sind. Keine Abschnitt der Doktorandarbeit wurde von einer anderen Person formuliert, und bei der Abfassung wurden keine anderen als die in der Abhandlung aufgeführten Hilfsmittel benutzt.

Is written by myself, I have formulated all parts independently and I have not taken any texts components of others without indicating them. No formulation has been made by someone else and I have not used any sources other than indicated in the thesis.

Ich erkläre, das ich habe an keiner anderen Stelle einen Antrag auf Zulassung zur Promotion gestellt oder bereits einen Doktortitel auf der Grundlage des vorgelegten Studienabschlusses erworben und mich auch nicht einer Doktorprüfung erfolglos unterzogen.

I assert I have not applied anywhere else for a doctoral degree nor have I obtained a doctor title on the basis of my present studies or failed a doctoral examination.

Ort, Datum

Place, date

Unterschrift Doktorandin/Doktorand

Signature of the doctoral candidate

Acknowledgments

The research presented in this work was supported by the Elite Network of Bavaria [Project number: K-GS-2012-209]. I would like to extend my sincere gratitude to the opportunities (conference attendance, incubator stay, international knowledge exchange) made possible by the ENB.

I would like to extend my sincere gratitude to the Munich Center of the Learning Sciences; the coordinators, secretaries, and professors of the REASON program. You helped us –The REASON students- not only to get through the administrative chaos, but also for showing us how to survive the academic environment, how to live in Munich, and even for helping us finding housing after a rough start.

This has been a long and rough journey, one that would not have been possible without the help of many people. During all these years, since I was 18 years old, I have suffered from depression. It is a hidden monster that rises when you do not expect it and like a *Dementor*, sucks away your happiness and hopes for the future. I have been fighting against this monster during all these years. Fortunately, I was not alone. I would not have been able to survive my studies and fight the monster just by myself, so the number of people to be grateful to goes beyond these pages. Finally, I arrived at the end of my PhD studies, the monster is still there, but I had fought against it and made it to the end despite it.

I want to thank my first supervisor, Professor Birgit Dorner. You believed in me and my project. You have been helpful in every aspect of my project from literature search to the search and directing of student assistants for the project. You were not only a great help but a great inspiration for my project and for future projects I would like to conduct and learning I would like to do.

I want to thank my second and third supervisors, Professor Heinrich Hussmann and Professor Jan-Willem Strijbos for being so supportive during the research process, not only with your academic guidance but also for believing in me despite the different type of research I was doing. I would not have been able to do it without your advice on technology, methodology, bibliography, style, and bureaucracy. You also spend a lot of your time helping me reviewing my writings and figuring out how to proceed in an uncertain terrain. All your help and techniques will surely be helpful for my future career. Thank you very much.

Thanks to my international supervisor, Professor Richard Bartle, for welcoming me as your supervisee. You are both a great person and a great professor, and your advice both for my studies and for my future career have already proven to be valuable. It is hard to find people so *savvy* and humble at the same time. I also want to thank the IGGI group at the University of Essex, especially Andrei Iacob, Mihail Moroșan, and Joseph Walton-Rivers for welcoming me into their team, helping me understanding the life at the UK, and catching up with years of idle programming skills.

I also want to extend my gratitude to Prof. Florian Alt for helping me understanding and entraining in the world of eye-tracking studies. Also to my students Marielle Dado and Laura Delonge for helping me gathering the data, organising the schedule, understand the game and the players, and sometimes make me reflect on my work finding new links.

I would like to thank Mario Alba, Liliana Chaves, Johny Villada, and Professor Alessandro Canossa for their ideas, insights, and help in my project. Although you did not get anything special for helping me, you gladly assisted me when I was lost in my research and when I had lost faith in my project.

I want to thank my father, José Alberto, who died two months before I started my PhD studies. I owe him my education at many levels, including the languages. Although he did not

live to see me achieving a doctoral degree, I think he will be proud of me for my achievements.

I warmly thank my mother Fabiola; you gave me everything and helped me in the most difficult times of my life as well as celebrating the most joyful ones. You made me who I am by teaching me the scientific road and allowing me to experiment, make mistakes, and analyse them since I was a child. This journey would have been impossible without your help; therefore I owe you everything I am.

Last, but not least, I thank my beloved husband Jesper. You have been there in the most difficult times, and have had to deal with difficult times yourself in order to support me. You have also been there to celebrate the nice times and happy days. You are the best, and I would not have been able to do it without your support and your love.

My biggest gratitude to all of you, named and not named, you know I would not have been able to achieve it without your support, your help, and your teachings.

To my beloved mother Fabiola and my amazing husband Jesper.

You are my everything!

"Two roads diverged in a wood, and I—

I took the one less travelled by,

And that has made all the difference."

- Robert Frost

Contents

Abstract

The present research aims to inquire how Multiplayer Online Battle Arenas (MOBAs) foster Scientific Reasoning. The work of research takes place in two phases. In the first phase, we conducted a concept review of the concepts of Scientific Reasoning (SR) and Scientific Thinking (ST) to operationalise the variables for the main study correctly. The concept review was conducted in the first phase as the concept of SR was found to be confused with the concept of ST. Therefore, there was a necessity for differentiating and characterise both constructs before starting the main research. In the concept review, three questions were tackled: (1) Are Reasoning and Thinking the same processes? (2) Are SR and ST the same construct? (3) How can SR and ST be characterised to improve their research in the future? To clarify this ambiguity, we conducted a conceptual review using an integrative approach (qualitative, quantitative, and network analyses) with 166 texts acquired from five databases and 18 different disciplines to critically analyse the concepts of SR and ST. We found that Thinking and Reasoning can be characterised as different processes. Likewise, ST and SR can be characterised as distinct concepts/constructs. The review identified recent studies by which SR and ST could be further clarified from an ontological and a teleological perspective.

After the concept of SR was revised and an operationalisation was proposed, we conducted the second phase of the study. In the second phase, we inquired about how MOBA players use SR when playing, and its relationship with Strategy Making (SM) in the game. This phase explores four questions: (1) How MOBA players use SR and SM when playing, (2) whether differences in the level of expertise of players describe patterns of SR and SM on their play, (3) whether there is a difference in both SR and SM according to the game played, and (4) the relationship between SR and SM in MOBA game players. This research uses a mixed-method design, with semi-structured interviews and eye-tracking technologies to determine the role of SR and SM during the gameplay. The interviews were used to understand how players use in-game information, make hypotheses about the game, and use these hypotheses to create strategies allowing them to master the game. The eye-tracking data was used to uncover players' information seeking and information processing patterns during gameplay. The main findings were: (1) Players used SR with the main purpose of creating and deploying strategies for advancing tactical positions in the game, and (2) games belonging to the same category, genre, and playing style do not necessarily allow for close transfer of skills as the cognitive demand of games change according to variations in their mechanics. Mixed methods are required for studying video games when accounting for the complex characteristics of gameplay in a holistic way.

Finally, we verified that the way for characterising and operationalising SR found in the first phase of the study is appropriate for operationalising SR in the domain of Game Studies. This way of defining SR also accounts for fine and rough grain operationalisation in different qualitative and quantitative methods used in research.

Keywords: Scientific Reasoning, Scientific Thinking, eSports, Multiplayer Online Battle Arenas, Concept Review, Strategy Making.

Chapter 1

Introduction

The use of video games has increased in the recent years (Entertainment Software Association ESA, 2016). As stated by the annual report of the ESA, in 2016 63% of people in the U.S. play video games, indicating that the average U.S. household owns at least one game-dedicated PC, console or Smartphone. With an extensive number of consoles, game genres, and game developers in the market it is important to inquire about their influence on people's lives and skills.

Given the expansion of the use of video games, more research is being done every day to understand the effects of their use. Some works emphasise the usage of video games by the players (Devís-Devís, Peiró-Velert, Beltrán-Carrillo, & Tomás, 2009; Hofferth & Sandberg, 2001). Some researchers have focused on their negative influence like violence and addiction (Anderson, 2003; Anderson & Dill, 2000; Engelhardt, Bartholow, Kerr, & Bushman, 2011; Ferguson, 2007; Ferguson, Coulson, & Barnett, 2011), while others have turned to see how this apparently negative influence can be beneficial for the player (Durkin & Barber, 2002; Gollwitzer & Melzer, 2012).

Traditionally, psychological studies in video games have focused on studying violence and addiction in games (Anderson, 2003; Anderson & Dill, 2000; Khang, Kim, & Kim, 2013; King, Delfabbro, & Griffiths, 2013). A different line of psychological studies in video games combines learning and cognitive theories, aiming to explore the world of the serious, therapeutic, and educational video games (Burak, Keylor, & Sweeney, 2007; Galarneau, 2005; Orvis, Horn, & Belanich, 2008). Regarding positive human interaction with video games, there is a large branch of Psychology and Game Studies dedicated to enquire about the impact of serious and educational games on players (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; McClarty et al., 2012). Nevertheless, there is little research focusing on studying learning and development of cognitive skills by using commercial (off-the-shelf) video games (Baniqued et al., 2013; Blumberg, Altschuler, Almonte, & Mileaf, 2013; Gee, 2005a, 2005c; Smith, Stibric, & Smithson, 2013).

Electronic Sports –eSports (Hamari & Sjöblom, 2015) are a broad category of games which include Real Time Strategy games, First Person Shooters, Fighting and Arcade Games, and Multiplayer Online Battle Arenas (MOBAs). eSports, as recent phenomena, eSports have not been highly explored inside the academic community. Nevertheless, more studies, particularly within the field of psychology are emerging, aiming to understand how players play and which skills are used when playing (Bonny, Castaneda, & Swanson, 2016; Pöllänen, 2014; Wagner, 2006).

Multiplayer Online Battle Arenas are gaining popularity amongst the video gamer population as a consequence of their rise as an eSport. The current eSports MOBA tournaments move thousands of people around the world (Popper, 2013), millionaire prizes for the winners (Miller, 2010), and a whole community of professional eAthletes who train every day to be the best at performing on these new type of sports (Tassi, 2012). Even some schools and colleges around the world are considering eSports within their curricula, creating scenarios where students can participate in the old gym class as an eAthlete (Maiberg, 2014; Tassi, 2015).

The present work aims to explore in depth the relationship between the gameplay of MOBAs and Scientific Reasoning (SR). This research is done with the purpose of understanding better the cognitive processes activated when people play them, as well as the possible benefits of playing MOBAs. The results of this work will not only lead to a better comprehension of the use SR in informal contexts but also to better understand how games, particularly MOBAs can serve as a way to train cognitive skills. The current research studies the MOBAs League of Legends (LoL) and Defence of the Ancients 2 (DotA2) for being popular games, free to access and free to play, and with a big gaming community and eSports followers.

Before researching on MOBAs and their relationship with SR, it is necessary to deepen in the concept and characteristics of SR. The first step in research is searching for bibliography about the phenomenon to be studied. Bibliographic research in quantitative studies serves the purpose of getting acquainted with the topic, characterise it, and operationalise it. Following this process, we started investigating the concept of SR, promptly finding some problems in its ontology and characterisation: (1) The concept of SR is found to be mixed with that of Scientific Thinking (ST), sometimes finding them both as synonyms in the academic bibliography and sometimes finding them defined and characterised differently. (2) Different authors characterise and operationalise SR in various ways; therefore the concept is not unified for its study.

As we found this impasse, we decided to conduct a concept review on SR and ST to correctly assess and operationalise SR before starting studying how players use this cognitive process when playing MOBAs.

1.1. Objectives

1.1.1. Main objective

The main objective of the present project is to explore how MOBA players make use of Scientific Reasoning skills in the process of playing.

1.1.2. Specific objectives

1. Develop a theoretical framework guiding the rationale of the study concerning SR characteristics in gaming contexts.
2. Characterise and operationalise the concept of SR for the research.
3. Explore the patterns and characteristics of SR and its relationship with SM in Expert MOBA players' of two different games (LoL and DotA2).

1.2. Theoretical Framework

1.2.1. Scientific Reasoning

SR is a blurrily defined concept. Its definition ranges from problem solving tasks involving hypotheses generation, testing, and strategic thinking (Newell & Simon, 1972); to a specific set of abilities required to understand the nature of science, theories, design experiments, and interpret data (Koerber, Sodian, Thoermer, & Nett, 2005).

The first step in research is searching for bibliography about the phenomena to be studied; this serves for getting acquainted with the topic, characterise it, and operationalise it. Following this process, we started investigating the concept of SR, promptly finding three problems in its ontology and characterisation. First, SR often gets mixed with the concept of ST raising the question if Reasoning and Thinking are the same activities. The second problem is if SR and ST are the same concept or are different but used imprecisely in the literature. The third problem is that not all authors define SR in the same way; some focus on a specific set of cognitive skills people develop, and others focus on constructs derived from the history of scientific enquiry and development.

These three problems derived in our fist study presented in Chapter 2. Aiming for a better characterisation and operationalisation of the concept of SR we conducted a conceptual review of the concepts of SR and ST striving to solve the three inconsistencies found.

1.2.2. Games, video games, development, and reasoning

First of all, it is necessary to define what play, game, and video games are. As stated by Walther (2003), there is an ontological and epistemological difference of play and game. According to Rogers (2010) and Walther (2003) playing is an activity that is auto motivated and which boundaries, goals, and rules are not defined; thus, playing provides enjoyment *per se* to its executor. Conversely, games are structured systems; they are closed, formal, based on rules, limited in space and time, and have an unequal goal which can be attained or failed (Fullerton, 2008). For Rogers (2010) the definition of video game corresponds to the same definition of game, but it is played on a video screen.

Huizinga (1950) in his book *Homo Ludens*, characterise play in a form that right now is known as game, but his characterisation of a "game" plays a fundamental role in Ludology studies. According to Huizinga, a game possesses six fundamental characteristics: (1) A game is voluntary and cannot be subject of imperatives; (2) it is pretended (it does not emulate the real life); (3) it is immersive (when someone plays a game, that person feels to be part of another world, losing the sensation of time and space around them); (4) it has boundaries (is limited in space and time); (5) it is based on rules; (6) people involved in it tend to identify themselves as a group.

The studies about the activity of playing and gaming made by Jean Piaget (Lipsit & Reese, 1981; Singer & Revenson, 1996) and Lev Vygotsky (Vygotsky, 1979) about the role of the play and the game in children development are well known within the field of

Psychology. These studies pose as the angular stone for understanding the relevance of playing and gaming in the human development.

Piaget (1962, as cited in Singer & Revenson, 1996) states that the playing activity starts from birth as an imitation process. This process is, in the beginning, mostly accommodation (i.e., fitting extraneous patterns to patterns or schemata they already possess), but it quickly escalates and develops into sophisticated ways of play. The role play, for instance, allows children child to explore and practice their mechanism of assimilation and accommodation; that is to say, to apply their schemata to real-life phenomena, to change their schemata and create new ones based on problems given by the real world while in a safe environment.

Piaget (1962, as cited in Singer & Revenson, 1996) asserts that there are three developmental stages of play which are in correspondence with the stage of development proposed by the author: Imitation, symbolic play, and game with rules. For Piaget, the imitation period goes from the zero to the two years of age of the child. It starts with simple reflexes and repetition of acts; developing into true imitation (i.e., imitation of parents behaviour), the inclusion of what they cannot see (language), and ritual behaviour. The most evolved behaviours of this phase include the imitation of complex behaviour and symbolic imitation (i.e., the parents' roles).

The second stage is the symbolic play; it emerges and develops between the two and seven years of age. This stage is characterised by make-believe games (e.g., the child can "be" *Spiderman*), distortion of reality (e.g., a broom can "be" a horse) and representation of imaginary objects (e.g., imaginary friends). Additionally, this phase is known for the autistic play (children tend to play alone in their imaginary world, even when they are with other children).

The final stage for Piaget is the game with rules; it develops between the seven and eleven years of age. Children normally start playing institutional (cultural) games like *hopscotch*. Later on, they start entering the adult world by playing more elaborated and ruled games, such as *chess*. For Piaget, this phase is when the child develops the psychological characteristics (adapt to and accept rules) allowing them to enter the adult world.

For Vygotsky (1979) the activity of playing is closely related to the development of the child within the Zone of Actual Development and the Zone of Proximal Development. Playing enables the child to practice their skills. At the same time, it allows children to explore, discover and learn how to do things they could not do alone before. In other words, it expands the Zone of Proximal Development

For Vygotsky (1979) playing is a social activity, linked to a socio-cultural context; comprising the people, language, social, and technological artefacts developed in a certain socio-historical context.

Like Piaget, Vygotsky (1979) also proposes some phases on the development of the playing activity of the child. First, there is the literal play, where the child cannot detach from the symbols/meanings present in real life (concrete operations stage for Piaget). Then the child starts using pivots, which allow them to change from concrete operations to symbolic operation (e.g., a broom can "be" a horse). Finally, the pivots become internalised, and the child can abstract play (symbolic operations stage in Piaget). Moreover, Vygotsky states that playing allows the children to develop social rules, as well as developing the self-regulation required to join the social world and more complex activities.

1.2.3. Serious games and commercial games

With the increasing usage of video games as an everyday leisure activity, some theorists have started to investigate their educational value (Gee, 2005a, 2006, 2007a, 2007b, Prensky, 2002, 2006). While some specialized video game producers have focused on learning and Serious Games created with educational and curriculum contents (de Castell & Jenson, 2007; Galarneau, 2005; Rosas et al., 2003; Rosas, Grau, Salinas, & Correa, 2000), commercial-non-educational video games continue to be the most sold and played ones (Helm, 2005). Concerning this, some researchers have proposed that commercial video games possess a learning value, and can be used in the classroom for learning purposes (Gee, 2005b, 2007a, 2007b, Prensky, 2005, 2006). Studies have been conducted to identify if video games increase the capacity of learning (Boot, Kramer, Simons, Fabiani, & Gratton, 2008; Dye, Green, & Bavelier, 2009; Gee, 2005b, 2005c; Juul, 2007; Prensky, 2006), or if a game developed for the classroom has the persuasive power to help children to learn or improve their cognitive skills (Burak et al., 2007; Nacke, Nacke, & Lindley, 2009; Rosas et al., 2003).

As defined by Michael and Chen (2006), Serious Games are games maintaining the essential definition of games, but which objective is to deliver educational content rather than entertainment. As stated by Abt (1987, as cited in Michael and Chen, 2006, p. 21) "[Serious Games] have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement". In this sense, Serious Games always violate at least one of the characteristics mentioned by Huizinga (1950) (i.e., they might not be voluntary when introduced to a classroom). Fun is not the goal of a Serious Game, but it is considered by most of the theorists on this topic as a side feature of Serious Games.

According to Michael and Chen (2006), there are different types of Serious Games according to their use, goal, and presentation:

1. *Educational video games* are games which content has been thought-out focusing on learning elements, these can be either for learning proficiencies (e.g., learn to read), or content (e.g., learning biology). Educational games can also be presented in the form of analogue games (e.g., board games), in which case they are not video games.

2. *Edutainment games* are games derived from edutainment contexts. These games are rigid compared to the educational games but offer a more scaffolded learning environment.
3. *Simulations* are often used for training real-life situations. Their aim is to replicate a real-life scenario as faithful as possible to the real scenario.
4. *Sensitization and information games* aim to familiarise the player with a situation (normally a social situation) the player does not know or do not understand. The goal of these games is to get the player closer to a phenomenon.

Additionally, Michael and Chen (2006) suggest that Serious Games can be used for training, learning and teaching, sensitise and inform, enhance work practices, healthcare, and as a means of art.

Nevertheless, James Paul Gee (see for example Gee, 2005c, 2006, 2007b), asserts that games in their "natural way" are always teaching something. For Gee, games and video games, in the same line as stated by Piaget and Vygotsky, represent a zone the player needs to access to get to the fun. To get to the fun, the player must take the challenge to learn a set of rules and procedures. For Gee, games are teaching instruments by themselves; moreover, fun is the key element for games to be successful in delivering educational content. Michael and Chen (2006, p. 26) indicate how "Immersion in simulated environments increases learning speed and retention for a range of tasks".

There are two positions concerning Serious Games, one focusing on delivering well-thought educational content, and the second one prioritising the fun and the immersion when trying to deliver educational content. The current research makes use of the second approach, stating that non-Serious Games (also known as commercial games, recreational video games, or off-the-shelf games) deliver educational content in the same way as stated by Gee (2005a).

Putting aside controversies surrounding the commercial and recreational video games such as addiction or vicarious learning of violence, it is undeniable that they possess ludic and recreational value. Gee (2005b, 2005c, 2006, 2007b) and Prensky (2006), have studied and advocated for the acknowledgement of video games beyond a leisure activity, characterising the beneficial effects of playing video games in different aspects of human development such as cognitive abilities, motor skills, and knowledge.

1.3. Description of the study

The present project was conducted in two phases; the first phase is elaborated in Chapter 2 and aims for a better understanding of the concept of SR for providing a characterisation and operationalisation to be used in the main study about MOBAs. This first step aimed to answer three Research Questions (RQ) related to the concept of SR: (RQ1) Are SR and ST the same concept or different concepts? (RQ2) If SR and ST are not the same concept, what are their differences? (RQ3) How can we characterise and operationalise the constructs of SR and ST so that they can be more systematically used in research? To clarify this ambiguity, we conducted a conceptual review using an integrative approach (qualitative, quantitative, and network analyses) with 166 texts acquired from five databases and 18 different disciplines to critically analyse the concepts of SR and ST.

The second phase is elaborated in Chapter 3, and it aims for understanding how players of LoL and DotA2 make use of SR during the gameplay and in which way it is presented during the gameplay. This phase aimed to explore four Research Questions concerning SR in MOBAs: (RQ4) How MOBA players use SR and Strategy Making (SM) when playing; SM was adopted in the research because it is closely related to SR and allows to uncover aspects of SR that are not directly observable. (RQ5) whether differences in the level of expertise of players describe patterns of SR and SM on their play; (RQ6) whether there is a difference in both SR and SM according to the game played; (RQ7) and the relationship between SR and SM in MOBA game players. This phase makes use of a mixed-method design, employing semi-structured interviews and eye-tracking technologies to determine the role of SR and SM during the gameplay. The interviews were used to understand how players use in-game information, make hypotheses about the game, and use these hypotheses to create strategies allowing them to master the game. The eye-tracking was used to uncover players' information seeking and information processing patterns during gameplay.

On Chapter 4 we raise a general discussion on the two phases of the research while accounting for the usability of the construct of SR developed during the first phase in Game Studies. Moreover, we discuss the theoretical and practical implications of the study and suggest some points for deepening in future research regarding the subjects of the two different phases of the project.

Finally, we summarise and reflect on the challenges of researching on commercial video games and propose a way to approach to their study based on our research experience, decisions made, and learning processes.

Chapter 2

Conceptual review on Scientific Reasoning and Scientific Thinking

As part of high order thinking processes, Scientific Reasoning (SR) and Scientific Thinking (ST) are concepts of great relevance for psychology and education disciplines (Kuhn, 2009). The relevance of these concepts reside in two levels; first an ontogenetic importance for developmental psychology (Zimmerman, 2007) reflected in the early curiosity of the human mind for understanding their surroundings (Jirout & Klahr, 2012; Loewenstein, 1994) and fostered and supported by the society and the education system (Abdullah & Shariff, 2008; Adey & Csapó, 2013). The second level is of historico-cultural importance for which science has developed amongst the story, not as a collection of facts, but as a collection of explored, refuted, and renewed hypotheses about how the world works (Gower, 1997; Popper, 1962).

Although the constructs of SR and ST have been explored and studied by many disciplines for a long time, we found that these concepts are not clearly defined by authors studying them. The confusion with these terms has different facets. Next, we will present three problems found when researching on SR and ST, followed by an example of how the problem was found in different contexts.

2.1. Scientific Reasoning or Scientific Thinking

The first problem found when conceptualising SR and ST is that they often switch conceptual reference as if they were synonyms. Example 1 illustrates this issue, the text was found in Science Direct under the keywords ‘Scientific Thinking’.

Example 1

> Science educators who write about the development of *scientific thinking* [emphasis added] skills emphasize the extent and complexity of what needs to develop if students are to become effective science learners (Duschl, 2008; Fortus et al., 2006; Kuhn & Pease, 2008)—a complexity and extent clearly reflected in K-12 science curriculum standards (National Research Council, 1996, 2007). A long-standing tradition among developmental psychologists who study *scientific reasoning* [emphasis added], however, has been to focus attention on a single reasoning strategy, the control-of-variables strategy, featured by Inhelder and Piaget (1958) in their now classic volume (for early reviews, see Neimark, 1975, or Keating, 1980; for contemporary ones Zimmerman, 2007, or Kuhn, 2002.). (Kuhn, Iordanou, Pease, & Wirkala, 2008, p. 435)

This example mixes SR and ST in a sense the text itself proposes as keywords for reading the text the concept of ST; nevertheless, the text switches concepts from SR to ST as if they were synonyms.

Example 2 illustrates how the concept talked about is ST, but the keyword under which the text was found is SR. This example is taken from Science Direct under the keyword ‘Scientific Reasoning’. The text proposes as SR and ST as keywords for reading, using the two concepts as synonyms

Example 2

> *Scientific thinking* [emphasis added] is defined as the application of the methods or principles of scientific inquiry to reasoning or problem-solving situations, and involves the skills implicated in generating, testing and revising theories, and in the case of fully developed skills, to reflect on the process of knowledge acquisition and

change (Koslowski, 1996; Kuhn & Franklin, 2006; Wilkening & Sodian, 2005). (Zimmerman, 2007, p. 173)

Based on the confusion like in this example we decided to explore the following three Research Questions (RQ): (RQ1) Are SR and ST the same concept or different concepts? (RQ2) If SR and ST are not the same concept, what are their differences? Finally, (RQ3) How can we characterise and operationalise the constructs of SR and ST so that they can be more systematically used in research?

2.2. Are Scientific Reasoning and Scientific Thinking the same concept?

Not all authors characterise SR and ST to be the same construct. Some authors state that SR and ST are the same concept, using them as synonyms (Example 1); others define SR and ST in the same way but with a different name (i.e., SR or ST as a cognitive process). Other authors define ST to be associated with the history and development of sciences as a domain. Additionally, some authors explicitly differentiate SR from ST regarding them as different constructs. Example 3 and Example 4 illustrate how some authors use different names for the same construct. Example 4 and Example 5 illustrate how ST is regarded as a different construct in two different works. Example 6 illustrates how some authors explicitly trace the difference between SR and ST.

Example 3:

Research into children’s *scientific thinking and reasoning* [emphasis added] is multifaceted and includes the investigation of the formation and revision of theories and the principles of scientific inquiry (see Zimmerman, 2007, for a comprehensive review). The ability to select the correct test of a hypothesis is an integral part of the *scientific reasoning* [emphasis added] process, which is conceived of as including hypothesis generation, evidence evaluation, and experimental design skills (Klahr & Dunbar, 1988). (Croker & Buchanan, 2011)

This example was found in Academic Search under the keyword ‘Scientific Reasoning’. It depicts how SR (under the synonym of ‘Scientific Thinking’) is depicted as a process of hypothesis generation, evidence evaluation, and experimental design.

Example 4:

> The elements of *scientific thinking* [emphasis added] are essentially the same as for any reflective thinking. It is by the increasing awareness of the safeguards that must be thrown around the successive steps in the thought process that science has made its thinking constantly more cautious. The following outline will present these elements and safeguards [‘safeguards’ omitted]: Purposeful observation, analysis-synthesis, selective recall, hypothesis, verification by inference and experiment. (Downing, 1928)

This example was found in JSTOR under the keyword ‘Scientific Thinking’. It depicts how ST is depicted as a process of observation, hypothesis generation, and evidence evaluation using experimentation. Note the similitude with Example 3 where the concept of SR is defined.

Example 5:

> The problems of scientific progress, of conceptual development, of 'external' influences on science, of the relationship between science and the social order, cannot be properly enquired into unless an adequate theory of *scientific thinking* [emphasis added] and its relation to its subject matter, as well as an adequate theory of society, has been developed. (Harre, 2004, p. 29)

This example was found in Science Direct under the keyword ‘Scientific Thinking’. Different from Example 2 and Example 3, Example 5 depicts ST not as a cognitive process, but as a socio-historical phenomenon involved in the development of sciences. Note the difference with Example 4 where the concept of ST is also defined.

Example 6:

> One major branch of research into children’s scientific thinking has been concerned with *scientific reasoning processes* [emphasis added]. Such studies have sought to describe how individual children form hypotheses, collect evidence, make inferences, and revise theories. The *process of scientific thinking* [emphasis added] has been described as depending on the coordinated search of at least two problem spaces: a space of evidence and a space of theories (Klahr & Dunbar, 1988). These are seen as mutually interactive, so that inferences about evidence can modify theories and inferences drawn from theories can influence how individuals seek out further evidence. (Crowley et al., 2001, p. 713)

This example was found in PsychINFO under the keyword ‘Scientific Thinking’. It depicts how SR (under the synonym of ‘Scientific Thinking’) is related to the study of the hypothetico-deductive process of individual construction of knowledge, while ST (under the name of ‘the process of scientific thinking’ is related to understanding Scientific Discovery from a socio-historical perspective.

2.3. Lack of a referential ground for characterising Scientific Reasoning and Scientific Thinking

Based on Examples 1 to 6, it is possible to assert that the concepts of SR and ST are ill characterised. Additionally, different authors characterise them in different ways and associate it with different scientific concepts and teleologies; for instance, some authors

(Example 7) focus associate the concept of SR with a set of cognitive skills people develop; other authors (Example 3) associate the concept with scientific enquiry.

Example 7:

> Underlying our research design is the assumption that the reasoning strategies of interest to us exist at a level of generality greater than any one specific content domain. This assumption does not imply the view that these strategies are applied independent of a subject's specific knowledge within a domain. To the contrary, the content of subjects' theories within domains is central to our analysis. We regard *scientific reasoning* [emphasis added] as entailing the coordination of existing theories with new evidence bearing on them or, in the language used by Klahr and Dunbar (1988), the coordination of two "problem spaces." It is this coordination process that undergoes development (Kuhn, 1989; Kuhn et al., 1988) and is the focus of the present research. (Kuhn, Schauble, & Garcia-Mila, 1992, p. 287)

This example was found in JSTOR under the keyword ‘Scientific Reasoning’. This example (contrasted with Example 2) depicts how authors characterise SR in different ways, pointing towards a lack of consensus in the characterisation of the construct.

Additionally, some authors base their characterisations and operationalisation on secondary authors (e.g., Faulkner, Joiner, Littleton, Miell, & Thompson, 2000) or cognitive tests (e.g., Acar & Patton, 2012). Other authors characterise the concept of SR or ST without depicting a previous theoretical affiliation. Example 8 illustrates this issue. This example is taken from Science Direct under the keyword ‘Scientific Thinking’.

Example 8

> On the other hand, nonutilitarian (theistic, artistic, or scientific), nonexploitative thinking responses seem to involve neuronal processes within the Hss [Homo Sapiens

> Sapiens] brain that are founded on an emotional regard for humans as being superior to the physical world. Nonutilitarian rationalizations are motivated by quests to experience rewarding emotions associated with recognitions that the human mind is indeed superior to the physical world through discoveries of understandings of nature's different secrets and meanings of relationships among them. (Patterson, 1994, p. 3323)

This text exposes a different characterisation of ST created by the authors as a non-utilitarian process motivated by the human drive for understanding the world. This definition is created by the same author, and it does not have a theoretical foundation on authors that have developed the concept in a similar way.

Summarising, based on the different examples, it is clear that there is not only a lack of consensus on the characteristics and definitions of SR and ST but the concepts themselves lack a common definition amongst authors and disciplines, thus appearing as diffused and sometimes mixed.

The present research is a conceptual review of SR and ST conducted using convergent mixed methods (Creswell, 2015) with an integrative review approach (Soares et al., 2014; Torraco, 2005). This work is structured as follows: First, the research methodology and material collection process are described followed by the data analysis and results; afterwards, discussions and conclusions on the results are drawn. Finally, directions for future research and studies on SR and ST are depicted.

2.4. Research methodology

2.4.1. Design

The present study is a conceptual review (Alonso, 2013; Toubes, Santos, Llosa, & Llomagno, 2006) using an integrative approach (Gallardo-Echenique, Marqués-Molías, Bullen, & Strijbos, 2015; Soares et al., 2014; Torraco, 2005) to critically analyse and synthesise the concepts of SR and ST. The integrative literature review is defined as a type of research assessing and critically analysing pieces of literature about a topic (in this case a concept), allowing for integrating and synthesising frameworks and perspectives on the topic (Torraco, 2005). Using this method for a conceptual review grants rigour and systematicity, based on the research of extensive arrays of sources while acknowledging conceptual and theoretical developments (Callahan, 2014).

2.4.2. Sampling method

Five databases were used for searching conceptualizations of SR and ST: Science Direct, EBSCO Host Academic Search, EBSCO Host PsycINFO, JSTOR, and Google Scholar. The databases were chosen for hosting a broad spectrum of sciences and disciplines containing both experimental and theoretical works regarding the concepts studied. It can be argued that the database PsycINFO does not belong to this category because it is specialised in Psychology. Nevertheless, we chose it for its relationship with the concepts of reasoning and thinking. To compensate this choice we included the database JSTOR, which contains texts in Philosophy of Sciences and pieces of traditional research dating from 1900.

Queries were conducted searching for texts under the full excerpt “Scientific Reasoning” or “Scientific Thinking” in either the title or the keywords. The inclusion criteria were book chapters, journal articles, and conference proceedings published in English between 1900 and 2014. English was chosen for being a broadly used language in academic publications. Additionally, the data was filtered so only available full texts were obtained.

The inclusion criteria were restricted to the concepts of SR or ST which semantic value is related to the human construction of knowledge. Therefore, other constructs like those derived from Artificial Intelligence were assessed and discarded in the second phase of data filtering.

After each search, the formula for sample size was used to determine how many texts should be analysed from each result.

$$SS = \frac{Z^2 * p * (1 - p)}{C^2}$$

Where confidence level *Z* = 90%, the Confidence Interval *C* = 10%, the population size *p* depending on the number of hits obtained in the databases according to the concept searched. We used a theoretical distribution of 50%. Table 1 depicts how many results were products of each search and how many texts were selected in each case.

Table 1

Selection of texts according to the database, the excerpt used, and the search results

Database	Excerpt	Search results	Texts selected
Science Direct	Scientific Reasoning	191	58
	Scientific Thinking	143	49
EBSCO Academic Search	Scientific Reasoning	116	50
	Scientific Thinking	62	34
EBSCO PsycINFO	Scientific Reasoning	140	47
	Scientific Thinking	81	38
JSTOR	Scientific Reasoning	15	14
	Scientific Thinking	27	20
Google Scholar	Scientific Reasoning	86	37
	Scientific Thinking	67	36
Overall total		928	383

2.4.3. Filter and selection process

The data selected was examined to identify selected texts that were present in several databases. Table 2 depicts how many texts from each category were retained after removal of duplicates, and how many duplicates were found.

Table 2

Chosen texts for each sampling excerpt after repeated texts were discarded

	Raw data	After removal of duplicated texts	Duplicated texts in Databases
Total results for Scientific Reasoning	545		
Total selected for Scientific Reasoning	206	150	56
Total results for Scientific Thinking	383		
Total selected for Scientific Thinking	177	153	24

Afterwards, texts were first filtered for detecting texts not fulfilling the inclusion criteria. Texts outside the period between 1900 and 2014, editor letters, article reviews, book reviews, and theses were not used in the analysis. Table 3 depicts detailed information of the texts found under each category and therefore discarded.

Table 3

Texts found not fulfilling the inclusion criteria according to the sampling excerpts

	Editorials and opinions	Reviews and text comments	Reports	Theses	Total after the first filter
Scientific Reasoning	7	6	9	3	125
Scientific Thinking	5	31	15	2	100

The sample was then filtered for detecting texts where the semantic value of SR or ST was not related to the human construction of knowledge. An example of this can be observed in Laskey (2008) paper dedicated to developing a computer language for first-order Bayesian knowledge bases. The SR concept here is called *computational scientific reasoning* and describes computer information processing based on the scientific method. The concept of SR in Laskey cannot be compared with the concepts of SR and ST that can be traced back from 1900. For this filter, the abstracts, titles, and keywords of journal articles and proceedings were read. For book chapters, a skim read was conducted to detect if the concept described was relevant for the study.

After this, 24 texts were dropped from the SR pool of texts leaving a total of 101 items to review. Thirty-five texts were dropped from the ST pool of texts, leaving a total of 65 items to review. After the SR and ST texts had been joined, a total of 166 items met the inclusion criteria and were used for further analyses. An extensive list of the texts examined can be found in Appendix D.

2.4.4. Procedure

For analysing the texts, a Microsoft Excel database was created so that fine-grained analysis of the selected texts could be done. Table 4 explains how the database was built.

Table 4

Partial description of the fields contained in the text analytic database and description of each field.

<table>
<tr><th>Field</th><th>Description</th></tr>
<tr><td>First Author</td><td>Last name of the first author of the text.</td></tr>
<tr><td>Year</td><td>Year of publication of the text.</td></tr>
<tr><td>Title</td><td>Title of the article, book chapter, or paper in proceedings.</td></tr>
<tr><td>Research type</td><td>Research type according to (Ato, López, & Benavente, 2013).
1. Theoretical research: Narrative revision or update on theoretical studies about a research topic. Systematic review of basic research without the use of statistic procedures for integration of studies. Quantitative systematic revisions and meta-analyses.
2. Instrumental research: Pieces of work aimed to analyse psychometric proprieties of psychological batteries and tests.
3. Methodological research: Pieces of work presenting new methodologies related to the advance in the areas of research design, measuring in research, research analysis, simulations, and critical revision of methodological procedures.
4. Empirical research: Pieces of work using manipulation of variables as well as the acquisition of original empirical data and basic research.
Figure 1 depicts a diagram of the levels used by the author.</td></tr>
<tr><td>Research design</td><td>According to Ato et al. (2013), Theoretical research and Empirical research can be divided into different types of designs:
• Theoretical research
1. Narrative: Revision or theoretical update in basic research. It is rigorous but subjective and does not have empirical value.</td></tr>
</table>

(continued)

Table 4 (continued)

Field	Description
Research design	2. Systematic review: Theoretical revision or update with a systematic assessment of data analyses which do not use statistical procedures to integrate the data. 3. Meta-analysis: Quantitative systematic reviews consolidating basic research using statistical methods to integrate studies. • Empirical research 1. Manipulative: Pieces of research aimed to analyse causal relationships between two or more variables (causal hypotheses). 2. Associative: Non-experimental research aiming for finding functional relations between variables (hypothesis covariation) • Descriptive: Pieces of research describing phenomena as they occur, without variable manipulation or comparison, prediction, or modelling.
Research level	According to Ato et al. (2013), Manipulative, Associative, and Descriptive designs can be further divided into different levels: • Manipulative 1. Experimental: Studies where at least one variable is being manipulated and participants are randomly assigned to the treatments. 2. Quasi-experimental: Studies where at least one variable is being manipulated but participants are not randomly assigned to the treatments. 3. Single case: Experimental design where the number of units analysed is one or a very reduced sample. • Associative 1. Comparative: Studies based on the comparison between groups or subjects. 2. Predictive: Studies based on the prediction of behaviour or group classification. 3. Explicative: Studies aiming to test a theoretical model for further theoretical development.

(continued)

Table 4 (continued)

Field	Description
Research level	• Descriptive 1. Observational: Studies where the objective is to register specific behaviours assigning them to arbitrary codes. 2. Selective: Studies where the objective is to register opinions or attitudes using a scale or survey.
Research sub-level	According to Ato et al. (2013), Experimental research can be categorised according to the type of treatment, the type of control used, and the type of error (sample size) into levels. In the present study, we refer to them as 'empirical sub-level'. 1. Case-controls: Studies aiming to identify variations between two groups in function of the dependent variable. 2. Pre-post evaluation (developmental): Studies aiming to identify changes happening over a time period due to an intervention or developmental age of the participant. 3. Cross-sectional: Studies conducted in a defined temporal situation used to evaluate inquiries about the prevalence of a phenomenon and where the change over time does not play a major role. 4. Cohort (prospective): Studies evaluating a group forward in time for a time period using one or more dependent variables. 5. Ex-post facto: Retrospective study aiming to analyse one or more dependent variables on a particular event or population back in history.
Variables operationalised	Description of how the variables were operationalised (only for experimental studies). NA is assigned if the study is not experimental.
Concept characteristics	Characteristics under which the author conceptualised SR or ST
Teleology	Ultimate goal of SR or ST as stated by the author. (e.g., Scientific reasoning is to reason as scientist do)
Conceptual interchange	Yes or No category. Yes is assigned if there is an exchange of terms from SR to ST in the text or vice-versa; or if the author uses the terms indistinctively.

(continued)

Table 4 (continued)

Field	Description
Associated concepts	Concepts introduced by the author and that are Associated to SR or ST, such as scientific argumentation or scientific method.
Research question	Research question of the study for experimental studies.
Domain	Domain in which the research was conducted.
Sample size	Only for experimental studies. Sample size of the experiment.
Context	Only for experimental studies. Characteristics of the population and sample studied.
Findings	Only for experimental studies. Findings of the research and answers to the research questions.

2.5. Data collection

2.5.1. Data management and data processing

After the selected publications had been decomposed into the database, an analytical procedure was conducted with the fields of "Concept Characteristics", "Teleology", and "Associated Concepts". These fields were narrative and could not be directly assessed for conceptual regularity. The analytical procedure was done so conceptual categories could be derived from and associated with the narrative description. The ideas from the authors were gathered in one to three words. For instance, Hsieh, Cui, and Sharma (2008) characterise ST as:

> "The ability to identify the problem at hand while attending to the related context, to support arguments with credible data, and to conclude the influence and impact of the problem at hand." (p. 124)

The last paragraph was analytically divided into the concepts of "Evidence Based, Problem Solving, Argumentation, Problem Identification, Context Sensitive". The list of concepts derived from the narratives for the “Concept Characteristics”, “Teleologies”, and “Associated Concepts” can be found in Appendix A.

The “Domain” field was expanded to match categories of “Main Domain”, and “Secondary Domain”. This division was done because some domains involved in the research were found to be specific sub-domains of another discipline (e.g., Education in Psychology is different from Educational Psychology; the first concerns teaching of Psychology, which main domain was categorised as Education and the secondary domain as Psychology. Educational Psychology, however, is a sub-domain of Psychology, making the main domain Psychology and the secondary domain Educational Psychology).

Afterwards, a special database was created in Microsoft Access to synthesise the fields of “Concept Characteristics”, “Teleology”, and “Associated Concepts”. This procedure was done because each field had more than one entry, which required the creation of a relational database allowing for the entry of multiple data in these fields. After the relational database had been finalised, it was exported back to Microsoft Excel.

Shorter versions of the main database were created focusing on the variables required to conduct a descriptive and network analyses for answering the research questions. Also, special dichotomous databases were created using R version 3.2.1 (the code used to create the databases and for analysing the data can be found at the repository from Díaz (2016)). These databases were created to assess individual components of the entries on fields with multiple answers. The dichotomous databases allow for creating descriptive statistics about the contents of the fields “Concept Characteristics”, “Teleology”, and “Associated Concepts”; as well as the creation of a network analysis database.

2.5.2. Validity and Reliability

Validity and reliability of the study were assessed following Akkerman, Admiraal, Brekelmans, and Oost (2008) audit methodology for evaluation research quality. This process accounts for objectivity and reliability during different stages of the research: (a) Designing and writing the research proposal, (b) gathering data, (c) data analysis, results and conclusions, (d) reporting the research. This process involves a dialogical interaction between the main researcher (first author) and an auditor (second author) in which documentation for each step of the process is created and discussed looking to improve the research in every stage. This process is iterative and finishes with a report from the auditor concerning the integrity of the work on each of the steps taken during the research. The auditor's report can be found in Appendix B.

2.6. Data Analysis

This conceptual review on the concepts of SR and ST was conducted using convergent mixed methods (Creswell, 2015) with an integrative review approach (Soares et al., 2014; Torraco, 2005). This section describes different types of analyses conducted with the sample. It is divided according to the three types of analyses, descriptive analyses to characterise the text sample, qualitative analyses derived from the comments and appreciations of the texts, and quantitative analyses aiming to deepen the results of the qualitative analyses. Finally, we conducted a network analysis aimed to understand differences and similarities between SR and ST based on their Characteristics, Teleologies, and Associated Concepts.

2.6.1. Descriptive statistics

One hundred and sixty-six texts were analysed for the study. Of the 166 texts, 35 were found in Science Direct, 46 in EBSCO Academic Search, 36 in EBSCO PsycINFO, 17 in JSTOR, and 32 in Google Scholar.

Regarding the resources included in the research, a total of 18 (10.84%) texts were book chapters, 19 (11.45%) were conference proceedings, and 129 (77.71%) were journal articles. From the total of texts, 65 (39.16%) were found using the keywords ‘Scientific Thinking’; and 101 (60.84%) were found using the keywords ‘Scientific Reasoning’.

The type of research for the sample was categorised using the levels proposed by Ato et al. (2013). Figure 1 shows a detailed distribution of the texts according to their appearance and the percentual representativity within the sample.

Figure 1. Distribution of works and representativity according to the research types proposed by Ato et al. (2013)

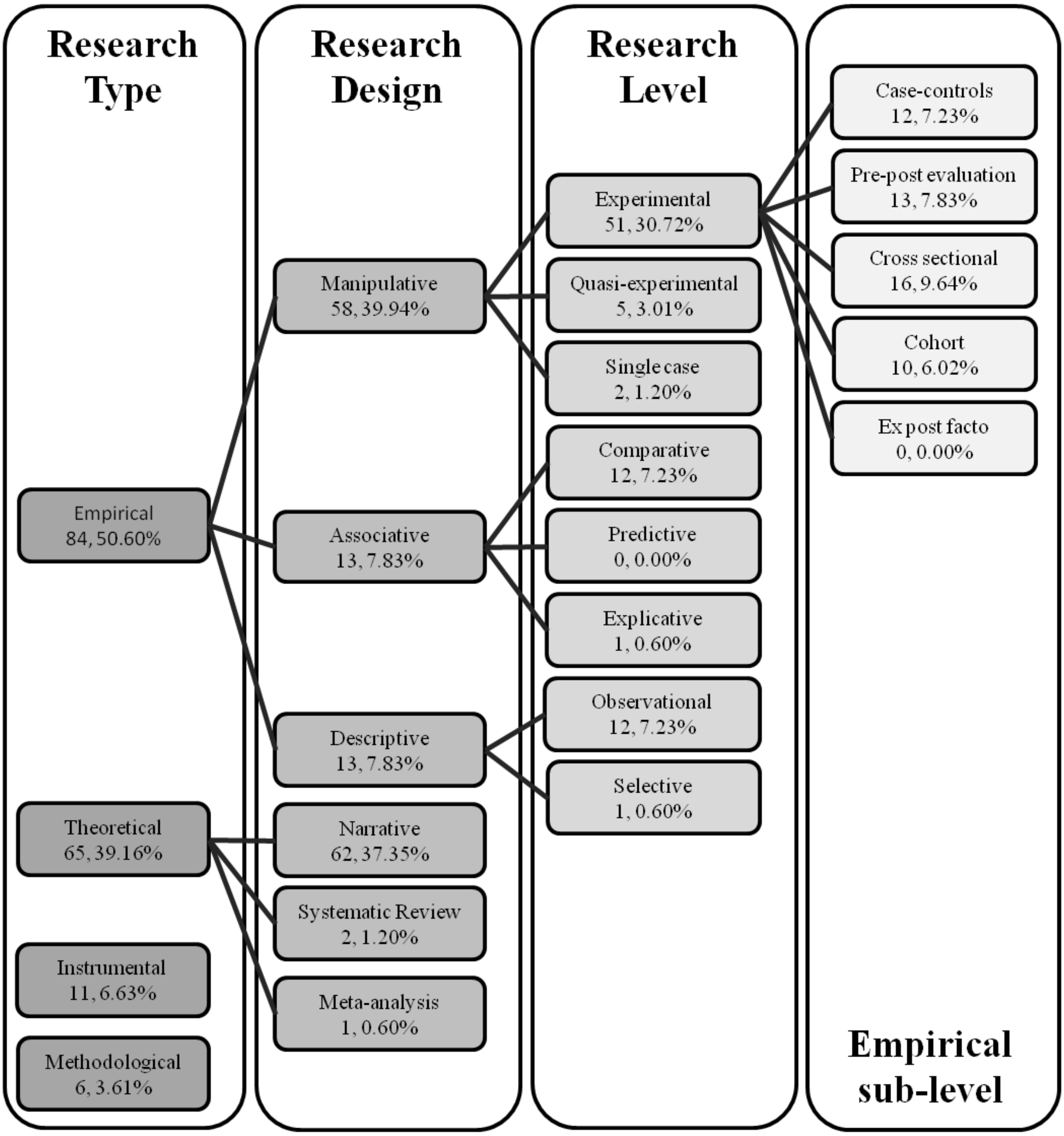


Figure 1. Distribution of the text sample. Under each name, the number of texts found and the percentage of representativity within the sample.

Additionally, frequency distributions for the Main Domain and Secondary Domain of study were conducted. Figure 2 and Figure 3 summarise the domains and their representativity in the sample.

Figure 2. Number of texts according to the Main Domain

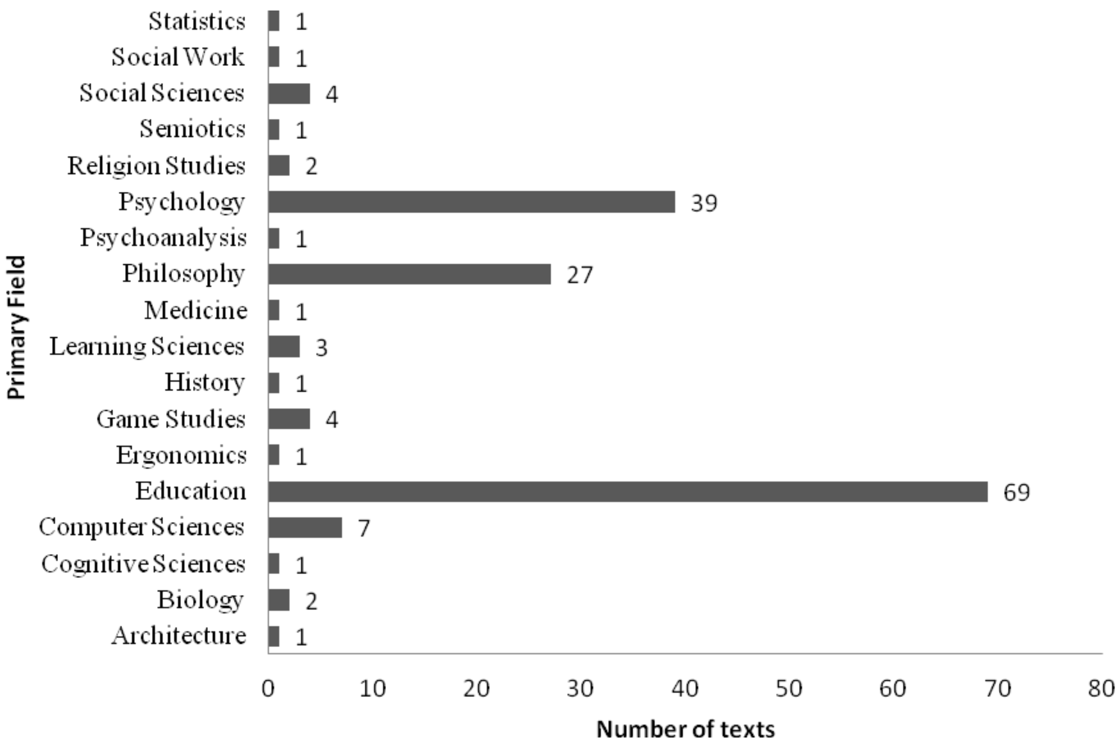


Figure 2. Eighteen Main Domains and their representation within the sample of texts. The domains of Education, Psychology, and Philosophy are the most salient in the sample.

Figure 3. Frequency of texts according to Secondary Domain

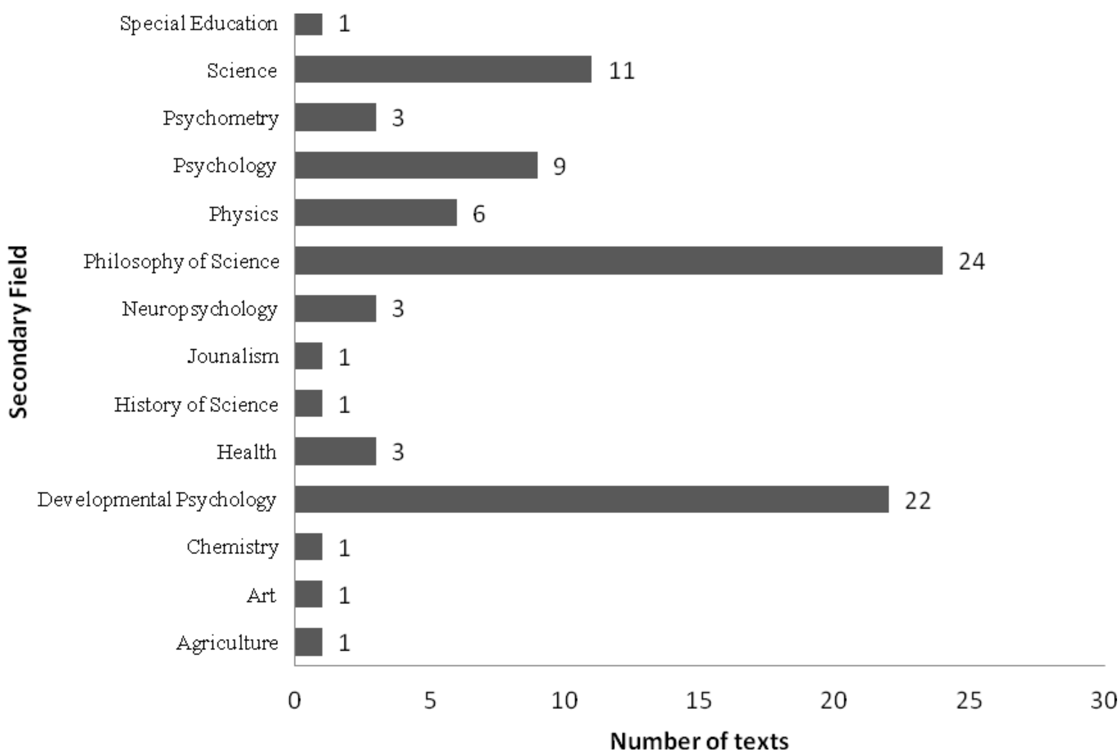


Figure 3. Fourteen Secondary domains and their representation within the sample of texts. The domains of Philosophy of Science and Developmental Psychology are the most salient in the sample.

The fields "Conceptual Characteristics", "Teleologies", and "Associated Concepts" were analysed for total frequencies for both SR and ST; these fields were analysed together accounting for the conceptual overlay between them -the differential analysis of the three fields was conducted using network analysis approach. For a better representation of the fields "Conceptual Characteristics", "Teleologies", and "Associated Concepts", only data above the average number of appearances are depicted. For an extensive list of terms found with their respective frequency of appearance, please refer to Appendix A. Figure 4 represents the most salient Conceptual Characteristics found in the data analysed, Figure 5 represents the most salient Teleologies found in the data analysed, and Figure 6 represents the most salient Associated Concepts found in the data analysed.

Figure 4. Distribution of the main conceptual characteristics found for SR and ST

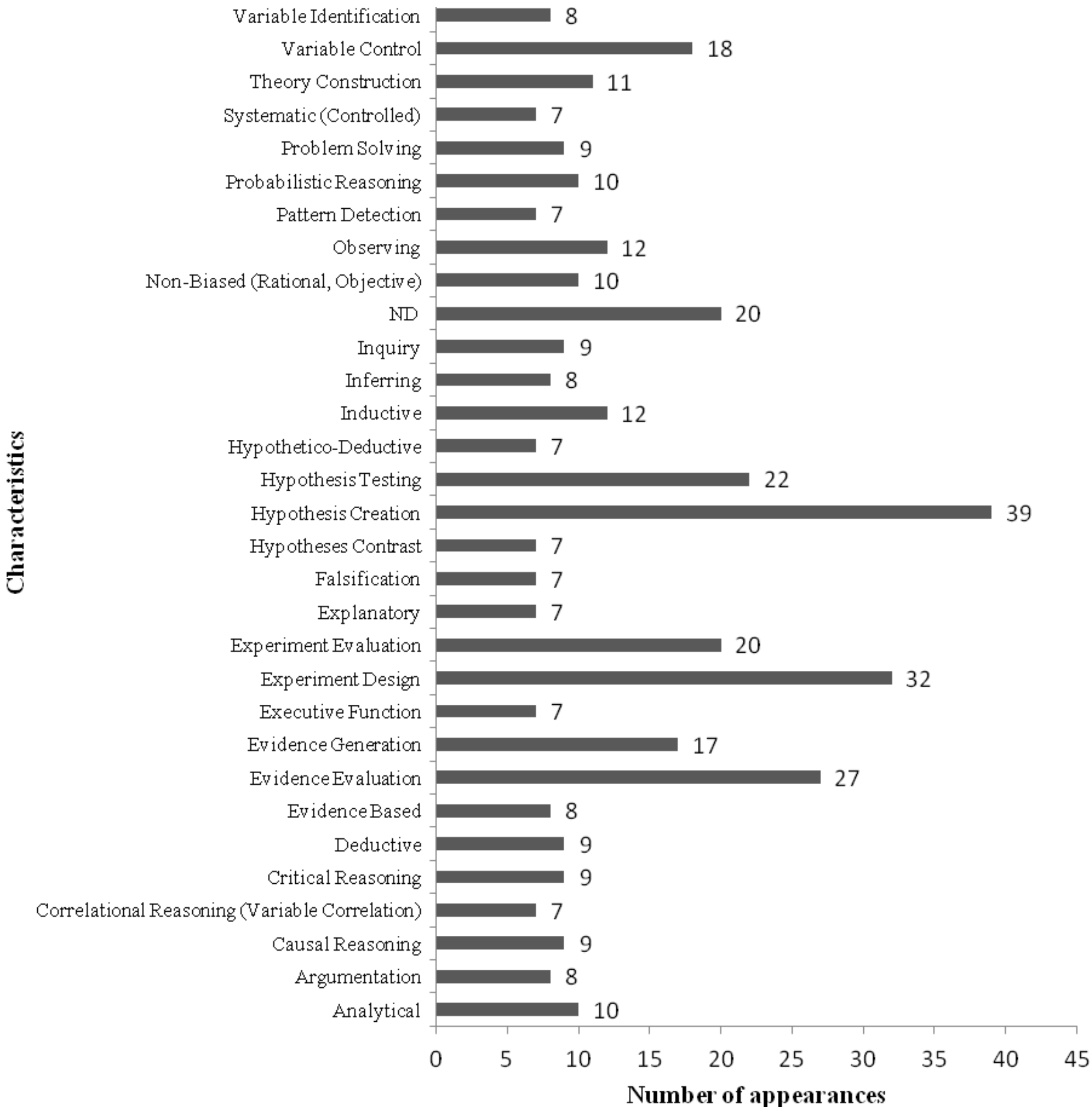


Figure 4. From a total of 102 characteristics, representation of the 31 most salients conceptual characteristics. Hypothesis Creation, Experimental Design, and Evidence Evaluation are the most repeated characteristics amongst the evaluated texts. ND does not correspond to a characteristic, but the number of texts not defining any characteristics for SR or ST.

Figure 5. Distribution of the main Teleologies found for SR and ST

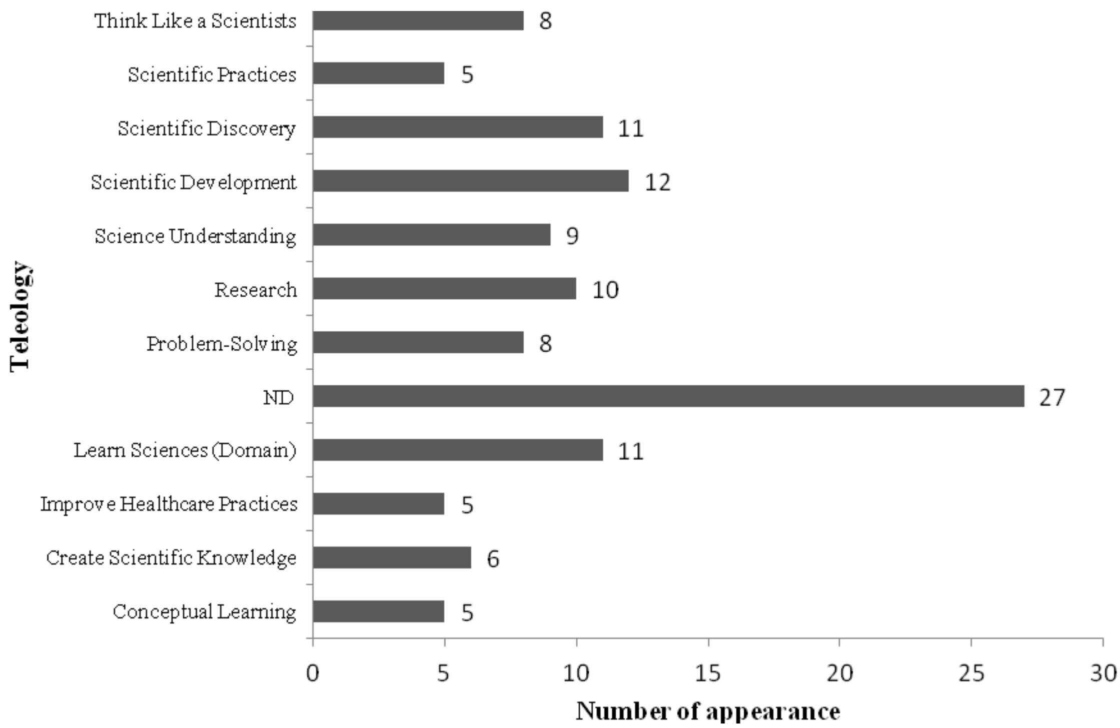


Figure 5. From a total of 52 Teleologies, representation of the 12 most salients. Texts with not defined characteristics were the most recurrent within the sample. ND does not correspond to a teleology, but the number of texts not defining any teleology for SR or ST.

Figure 6. Distribution of the main associated concepts found for SR and ST

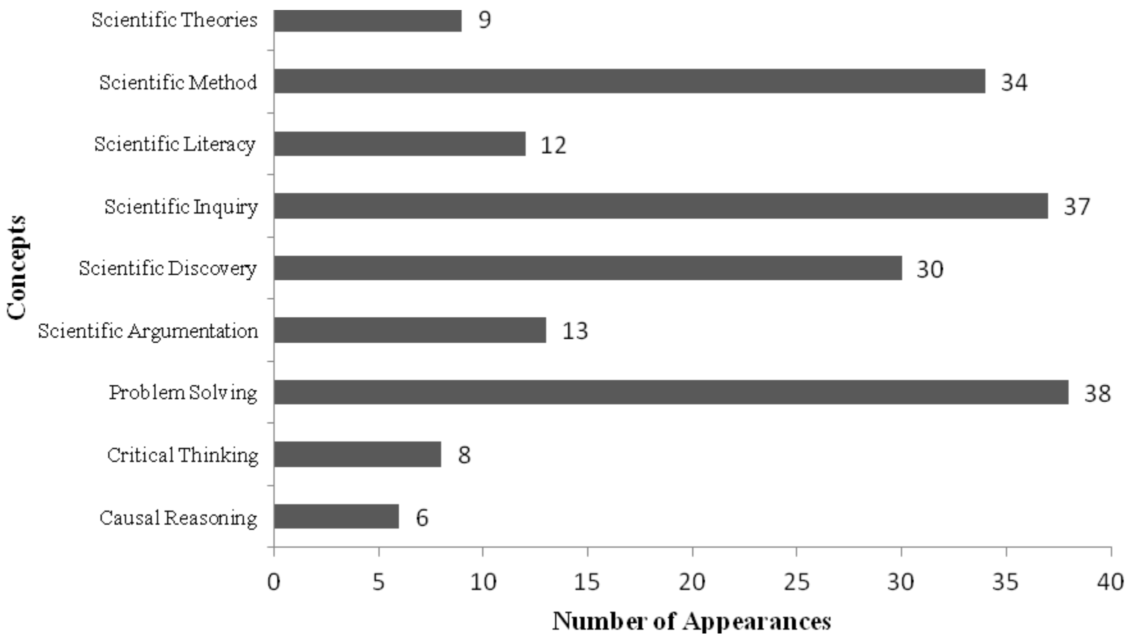


Figure 6. From a total of 51 associated concepts, representation of the nine most salients. Problem Solving, Scientific Inquiry, Scientific Method, and Scientific Discovery are the most repeated concepts amongst the evaluated texts.

2.6.2. Qualitative analyses

For the qualitative analysis, we made use of a summarising technique (Mayring, 2014) by consolidating the findings gathered while accounting for the perspective of the reader concerning the texts.

The qualitative data analysis revealed that some authors do not characterise the concept of SR or ST. Instead, they depart from a general perspective implying that the reader is already familiar with the concept of SR or ST and its characteristics (e.g., Faulkner, Joiner, Littleton, Miell, & Thompson, 2000). Other authors make a second level characterisation of the concept. That is to say, they mention other authors who have characterised the concept before them, but never characterise the concept in their research (e.g., Linn and Rice (1979) define the concept of SR based on the ‘Bending Rods task’ proposed by Inhelder and Piaget (1958), but without explicitly stating each of the characteristics evaluated). Both practices for characterising SR and ST are challenging for understanding empirical studies where concepts should be systematically characterised so they can be operationalised.

The second finding is related to the first. Some authors, particularly of experimental studies do not explicitly operationalise the variables they are researching or make a second level operationalisation based on tests (examples of this can be found in Abdulkarim & Al Jadiri, 2012a; Acar & Patton, 2012; Coletta, Jeffrey, & Steinert, 2011; Magno, 2011; Pyper, 2012).

A third qualitative finding is that some authors do not explicitly attribute a teleology to the concept (e.g., Azmitia & Montgomery, 1993; Beatty & Thompson, 2012; Hopkins, 1996; Linn & Rice, 1979; Tran et al., 2012). Other authors (i.e., Patterson, 1994) specifically state that ST is non-teleological. Additionally, we found intervention studies (pre-post or case-control studies) which did not state a teleology for their intervention (e.g., Collings, 1985; Gleason & Schauble, 1999; Siegler & Atlas, 1976). These findings raised new questions about the nature of SR and ST and their representation within academic circles. Potential directions for future research are raised in the discussion section.

The fourth finding was that some authors mentioned both SR and ST in their texts, but made the difference between the two explicit. These texts were coded as “no conceptual interchange” for not using SR and ST as synonyms but instead clarifying their difference; as an example, Crowley et al. (2001) consider SR and ST to be different, where SR is a sub-process of ST. SR is the process of making hypotheses, collect evidence, make inferences, and revise theories; while ST refers to the dialectics between theory and practice.

One last qualitative finding was the relation between problem solving and SR, thus linking the domains of SR, problem solving (Newell & Simon, 1972), and decision making (Kahnemann & Tversky, 1979) - Sinclair (1978) exemplifies this. The relationship between these domains proves to be insightful when trying to understand the characteristics and teleologies of SR as well as the conceptual difference between SR and ST. This relationship will be further explored in the subsection where we apply network analyses to infer patterns.

2.6.3. Quantitative analyses

Quantitative descriptive analyses were conducted for statistically supporting qualitative findings. Quantitative analyses were based on finding frequencies and proportions to support the qualitative findings. The first analysis conducted was regarding the number of authors not characterising the concepts of SR and ST or doing a second level characterisation. This procedure was done considering the main domain of study, the keyword used in the search, and the type of research. The results support the first qualitative findings and are summarised in Table 5.

Table 5

Non characterised texts by domain and research type for SR and ST

Main Domain	Research Type	Indetermination	Frequency	Percentage
Education	Empirical	External Definition	1	0.60%
		ND	7	4.22%
	Instrumental	ND	1	0.60%
	Theoretical	ND	3	1.81%
Game Studies	Theoretical	ND	1	0.60%
Philosophy	Theoretical	ND	6	3.61%
Psychology	Empirical	ND	1	0.60%
	Instrumental	External Definition	1	0.60%
	Theoretical	ND	1	0.60%
Social Sciences	Empirical	External Definition	1	0.60%

The second analysis reviews how many authors do not explicitly operationalise their research variable or rough-operationalise them based on standardised tests. It was found that most of the theoretical works of research (37.95% of the sample) do not operationalise the concepts, which is understandable because the aim of theoretical works is not to measure but to discuss historico-epistemological issues about the concepts. In contrast, it was found that 18.68% of Empirical research analysed in the sample do not operationalise their research or make a second level operationalization by stating that SR or ST will be measured with a standardised test. These results support the second qualitative findings and are summarised in Table 6.

Table 6

Non operationalised texts by domain and research type including both SR and ST

Main Domain	Research Type	Indetermination	Frequency	Percentage
Architecture	Theoretical	ND	1	0.60%
Biology	Empirical	ND	1	0.60%
	Theoretical	ND	1	0.60%
Computer Science	Empirical	Test	1	0.60%
	Methodological	ND	2	1.20%
Education	Empirical	ND	3	1.81%
		Test	16	9.64%
	Instrumental	ND	1	0.60%
	Methodological	ND	2	1.20%
	Theoretical	ND	11	6.63%
Ergonomics	Theoretical	ND	1	0.60%
Game Studies	Empirical	Test	1	0.60%
	Theoretical	ND	2	1.20%
History	Theoretical	ND	1	0.60%
Learning Sciences	Empirical	Test	2	1.20%
	Theoretical	ND	1	0.60%
Philosophy	Theoretical	ND	27	16.27%
Psychoanalysis	Theoretical	ND	1	0.60%
Psychology	Empirical	ND	4	2.41%
		Test	1	0.60%
	Instrumental	Test	1	0.60%
	Theoretical	ND	13	7.83%
Religion Studies	Theoretical	ND	2	1.20%
Social Sciences	Empirical	Test	2	1.20%
	Instrumental	ND	1	0.60%
Social Work	Theoretical	ND	1	0.60%
Statistics	Theoretical	ND	1	0.60%

The third analysis explores the relationship between domains and research types explicitly for the inclusion of teleologies in the study. This analysis includes both the concept of SR and ST. The results support the third qualitative finding and are summarised in Table 7.

Table 7

Non-defined teleologies by domain and research types for SR and ST

Main Domain	Research Type	Indetermination	Frequency	Percentage
Computer Science	Empirical	ND	1	0.60%
	Methodological	ND	2	1.20%
Education	Empirical	ND	4	2.41%
	Instrumental	ND	1	0.60%
	Theoretical	ND	2	1.20%
Game Studies	Empirical	ND	1	0.60%
History	Theoretical	ND	1	0.60%
Learning Sciences	Empirical	ND	1	0.60%
Philosophy	Theoretical	ND	3	1.81%
Psychoanalysis	Theoretical	ND	1	0.60%
Psychology	Empirical	ND	4	2.41%
	Instrumental	ND	2	1.20%
	Theoretical	ND	3	1.81%
Social Sciences	Empirical	ND	1	0.60%

A fourth analysis was performed to understand the amount of conceptual interchange of SR and ST in texts. A total of 111 texts were found where there was no conceptual interchange between SR and ST. Moreover, a total of six texts were found where there was no interchange and the differences between SR and ST are explicitly stated and explained (Crowley et al., 2001; Dejonckheere, van de Keere, & Mestdagh, 2010; Dunbar, 2001; Dunbar & Klahr, 2012; Kisiel, Rowe, Vartabedian, & Kopczak, 2012; Ruphy, 2011). These results support the fourth qualitative findings and are summarised in Table 8.

Table 8

Relationship between the concept explored (SR or ST) and the conceptual interchange inside the texts

Keyword Used	Conceptual Interchange	Frequency	Percentage
Scientific Reasoning	No	71	70%
Scientific Reasoning	Yes	30	30%
Scientific Thinking	No	47	72%
Scientific Thinking	Yes	18	28%

2.6.4. Network analyses

To explore the relationship between the SR, ST, and the different Characteristics, Teleologies, and Associated Concepts we conducted a network analysis which allows for identifying interlinked and independent concepts. The analyses were conducted using the package igraph v1.0.1 for R v3.2.1. Tables were generated based on the Network graph analysis to ease the reading and comprehension of the associated concepts. The Network Graphs can be accessed in Appendix C.

The first analysis comprises the relationship and differences between SR and ST based on the Characteristics described in the texts. Table 9 summarises the bonds between the concepts using their respective Characteristics. Afterwards, a qualitative summary (Mayring, 2014) was conducted.

Table 9

Network analysis for the concepts of SR and ST based on their characteristics

Scientific Reasoning only	Scientific Thinking only	Shared by SR and ST	
Bayesian	Auxiliary hypotheses	Abductive reasoning	Abstraction
Biased by beliefs	Context sensitive	Analogical reasoning	Analytical reasoning
Categories (Taxonomies)	Curiosity	Argumentation	Auxiliary explanations
Concept integration	Data interpretation	Causal reasoning	Classification
Confirmation	Distributed reasoning	Combinatorial reasoning	Communication
Conservation	Domain general	Complex	Concept generation
Discovery	Epistemology	Correlational reasoning (variable correlation)	Creativity
Domain specific	Generation of scientific laws and principles	Critical reasoning	Decision-making
Drawing conclusions	Non-Utilitarian	Deductive	Empirical evidence
Effect size	Operationalisation	Evidence based	Evidence evaluation
External definition (Non defined)	Processual	Evidence generation	Executive function
Extrapolation	Replication	Experiment design	Experiment evaluation
Generalisation	Sceptic	Explanatory	Exploratory
Heuristic	Strategy	Falsification	Hypotheses contrast
Historical	Structural	Hypothesis creation	Hypothesis testing

(continued)

Table 9 (continued)

Scientific Reasoning only	Scientific Thinking only	Shared by SR and ST	
Interpretation	Synthesis	Hypothetico-Deductive	Inductive
Intervention based	Understanding	Inferring	Inquiry
Mathematical		Knowledge generation	Logical thinking
Model revision		Measuring	Metacognition
Pattern detection		Modelisation	ND (not defined)
Rule creation		Non-Biased (rational, objective)	Not a collection of facts
Scientific writing		Observing	Prediction
Social		Probabilistic reasoning	Problem identification
		Problem solving	Proportional reasoning
		Reliability	Representational ability
		Science (as a Domain)	Scientific Method
		Separate scientific from non-scientific data	Statistical reasoning
		Systematic (controlled)	Theory-evidence coordination
		Theory construction	Theory revision
		Thinking scientifically	Variable control
		Variable identification	Variable isolation

The characteristics found to be exclusive of SR are, in general, closer to specific steps in the process of scientific inquiry like conservation of variables, and specific cognitive processes such as statistical reasoning. Characteristics found to be exclusive for ST are mainly related to general cognitive processes like pattern detection and general science topics.

The second analysis comprises the relationship and differences between SR and ST based on the Teleologies described by the authors of the texts analysed. Table 10 summarises the bonds between the concepts using their respective Teleologies.

Table 10

Network analysis for the concepts of SR and ST based on their Teleologies

Scientific Reasoning only	Scientific Thinking only	Shared by SR and ST
Argumentation	Access scientific knowledge	Coordinate theory and evidence
Conceptual development	Better represent the world	Create scientific knowledge
Conceptual learning	Develop a scientific attitude	Decision-making
Correct errors	Distinguish what is science from what is not science (e.g., science from pseudoscience)	Develop high order thinking
Develop critical thinking	Evidence-based practices	Evidence evaluation
Develop scientific artefacts	Improve game design	Hypothesis testing
Develop scientific understanding	Improve healthcare practices	Learn sciences (e.g., natural sciences and social sciences)
Discover laws and theories	Inquiry	ND (not defined)
Learn specific science (e.g., learn physics)	Know the truth	Predictions
Learn to think	Non-Teleological	Problem solving
Perform better in school	Understand human behaviour	Reason scientifically
Scientific literacy		Research
Scientific method		Science understanding
Scientific practices		Scientific communities
Survive the scientific publication world		Scientific development
Systematically tackle problems		Scientific discovery
Understand and solve social issues		Theory development
Understand covariation and correlation		Think like a scientist (e.g., a social or a natural scientist)
Understand the scientific method		Understand and solve scientific issues
		Understand complex phenomena
		Understand the world

Teleologies related to SR are highly specific regarding the development of cognitive skills such as 'understanding covariation and correlation', and the 'use and understanding of the scientific method'. In contrast, ST has two sets of teleologies. The first set is domain specific, such as 'improve game design' or 'improve healthcare practices'. The second set is related to the development of sciences in a socio-historical way such as 'better represent the world', 'access scientific knowledge' and 'understand human behaviour'.

The third analysis comprises the relationship and differences between SR and ST based on their associated concepts. Table 11 summarises the relationships between SR and ST using their associated concepts.

Table 11

Network analysis for the concepts of SR and ST based on their associated concepts

Scientific Reasoning only	Scientific Thinking only	Shared by SR and ST
Analogical reasoning	Belief based	Analytical reasoning
Decision-making	Collaboration	Causal reasoning
Dialogical	Determinism	Conceptual
Explanatory	Forecasting (Prediction)	Creativity
Falsification	Metaphoric thinking	Critical thinking
Formal reasoning	Paradigm creation	Epistemology
Integrative	Scientific achievement	Experimentation
Measurement	Scientific attitude	Exploratory
Multi-causal reasoning	Scientific discourse	Hypothetico-deductive reasoning
Science laws	Scientific invention	Modelisation
Scientific communication	Theory of Science	Problem solving
Scientific communities creation	Uses probes	Research
Scientific induction		Scientific argumentation
Scientific inference		Scientific discovery
Strategic		Scientific inquiry
Styles of scientific reasoning		Scientific knowledge
Understanding phenomena		Scientific literacy
		Scientific method
		Scientific practice
		Scientific revolution
		Scientific skills
		Scientific theories

Concepts associated to SR are mainly related to particular scientific achievements inside the academic community, like the creation of scientific communities, falsification of hypotheses and creation and deployment of experiments. Concepts associated with ST are related to the development of sciences as a socio-historical domain, focusing on the scientific discourse, the way scientists have developed methods and theories, the creation of predictive models, and bridging knowledge using metaphors.

2.7. Results

In this section, we combine and evaluate the findings made in the Data Analysis section. We start by integrating the descriptive findings with the research questions stated at the end of the introductory section. Next, we incorporate the qualitative findings with the frequencies supporting qualitative findings, thus presenting a holistic reading of the conceptualisation made by the authors of the works included in our selection. Finally, we explore the network analyses and its relevance for understanding the construction of the SR and ST concepts as well as the qualitative and quantitative findings.

The first part we want to highlight is that most of the sample we took is constituted by Empirical Research (50.60%) with an emphasis on Manipulative Research (39.94%), and Theoretical Research (39.16%) with emphasis on Narrative Research (37.35%) (Figure 1). This indicates a clear division between experimental studies (manipulative) aiming to test for the development and improvement of specific abilities related to SR/ST and theoretical reviews of the concept and its historico-philosophical nature. Interestingly, Instrumental and Methodological Research occupy a small part of the sample, indicating that most of the current Empirical Research on SR relies on instruments and methods previously constructed.

Out of the 18 main domains that we identified, most of the texts concerned the domain of Education (41.57%), followed by Psychology (23.49%) and Philosophy (16.27%) (Figure 2), marking a clear trend on the domains dedicated to inquiring about SR/ST. Concerning the secondary domains, out of the 14 domains found the most prevalent were Philosophy of Sciences (14.46%), Developmental Psychology (13.25%), and Education in Sciences (6.63%) and Psychology (5.42%) (Figure 3), supporting and defining a trend of research. The philosophical domains were dedicated entirely to Narrative Research (e.g., Popper, 1962), whereas the Education and Psychology domains were mostly dedicated to Empirical Research (e.g., D'Mello, Lehman, Pekrun, & Graesser, 2014) except for a few texts dedicated to synthesise (e.g., Cocking, Mestre, & Brown, 2000) or inquire (e.g., Kuhn, 2009) the current state of the research on the topic.

Characteristics regarding SR and ST that were most frequently mentioned in the texts are Hypothesis Creation, Experiment Design, Evidence Evaluation, Hypothesis Testing, and Experiment Evaluation (Figure 4). Pieces of work not defining the characteristics of SR or ST were found to be also frequent. The set of characteristics found is strongly connected to what is regarded in most of the texts as the process of SR, which involves an application of the Scientific Method: "A method of procedure that has characterized natural science since the 17th century, consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses" ('Scientific Method', 2016). Here, a person creates a hypothesis they want to test and then design an experiment where they contrast the fitness of their hypothesis with the evidence generated. The experiment evaluation appears here as an additional category of importance that is not directly related to the Scientific Method, and it is to revise the adequateness of the experiment for assessing if the results were fit and valid for solving the research question (Lilienfeld, Ammirati, & David, 2012).

Additionally, we found that some authors (23 authors – 13.86% of the sample- as seen in Table 5) do not characterise what SR or ST is, which is challenging when conducting empirical studies (e.g., McCabe & Castel, 2008).

Concerning the Teleologies found in the study, many of the authors do not define a teleology for SR or ST (Figure 5). This consideration is noteworthy in the case of empirical studies as SR or ST is not given a specific value regarding its goal (e.g., "why is it good to learn to reason scientifically?" or "how learning to reason scientifically can help a person in their lives?"). Our hypothesis is that for some authors the teleology of SR or ST is implicit, and it is not necessary to make it explicit. Next, the teleologies of Scientific Development, Learning Sciences, Scientific Discovery, and Research were identified as the most salient teleologies in the texts. In general, people reasoning scientifically develop the scientific domain, make scientific discoveries, and research. These are goals for both the education and scientific domains (e.g., Yeh & She, 2010), as well as advancing in sciences as a socio-historical construct (e.g., Schauble, Glaser, Raghavan, & Reiner, 1991). Learning (Natural) Sciences is a more specific goal found to be proper of primary education (e.g., Fender & Crowley, 2007), secondary education, and early higher education (e.g., Lazonder, Hagemans, & de Jong, 2010).

Regarding the Associated Concepts, we found Problem Solving, Scientific Inquiry, Scientific Method, and Scientific Discovery to be the most salient concepts in the texts (Figure 6). Problem solving was not only a recurrent topic in the quantitative findings but also in the qualitative findings. Moreover, some authors even state that SR or ST are the same as problem solving (e.g., Abdullah & Shariff, 2008; Azmitia & Montgomery, 1993; Klahr & Dunbar, 1988). The Associated Concept 'Scientific Method' was also highly present, sometimes being used as a synonym for SR/ST (e.g., Piraksa, Srisawasdi, & Koul, 2014; Thoron & Myers, 2012). Teleologies and characteristics such as "think like a scientist" were also related to this concept (e.g., Azarpira et al., 2012), raising the question of whether or not there is a distinction between SR/ST and the Scientific Method. This issue will be more extensively reflected upon in the discussion section. Scientific inquiry and scientific discovery are closely related concepts in the literature. The first is related to observation and a way of questioning what is observed (Holmes & Bonn, 2013). The second refers to finding patterns that help in understanding a phenomenon (Schauble et al., 1991), and in some cases to further develop a scientific domain (Dunbar, 2000).

Concerning the qualitative findings, we focused on five findings. First, authors either do not characterise the concept of SR/ST or make a second level characterisation. This finding was also supported by the descriptive data (Figure 4). To further investigate this finding, we performed additional quantitative analyses (Table 5). We found that among the main domains of study Education (7.23%) and Philosophy (3.61%) did not define characteristics. The higher number of undetermined entries by type of research were Theoretical (6.62%) and Empirical (4.82%). Both amounts of undetermined entries can be explained by the subsample size regarding Theoretical and Empirical studies.

Second, it was found that some authors of Empirical studies do not operationalise their research variables or make a second level operationalisation. To inquire about the lack of operationalisation, we made frequency analysis of domains and research types (Table 6). For the present analysis, we focus on Empirical pieces of work for being the ones that forcibly are measuring and comparing variables. A total of 4.21% of Empirical pieces of research within our sample were found not to be explicitly operationalised, while a total of 12.04% of the texts analysed depicted a second-level operationalisation, meaning that they did not directly operationalise the variables measured, but instead applied tests that theoretically measured the variables related to the construct of study.

Third, some authors do not explicitly attribute a teleology to the concept of SR/ST. This supports our findings from the descriptive analyses shown in Figure 5.

Fourth, we analysed the relationship between SR/ST and problem solving. First of all, problem solving it is one of the most salient concepts associated with SR (Figure 6). SR/ST and problem solving are associated in a twofold manner: Some authors state that problem solving is the underlying ability supporting SR/ST (Abdullah & Shariff, 2008; Azmitia &

Montgomery, 1993; Klahr & Dunbar, 1988), whereas others state that SR is the underlying mechanism behind problem solving strategies (Blair & Goodson, 1939; Nichol, Szymczyk, & Hutchinson, 2014).

Fifth, we conducted an analysis of the percentage of conceptual shift between SR and ST in texts (Table 8). The percentages of switching from one concept to the other are almost the same for both terms, pointing towards an almost indistinctive use of the terms for most of the disciplines analysed. One-third of the data analysed for both SR and ST can be considered to be ill-defined, as the concept switches in the text as if the concepts were synonymous.

Concerning the network analyses, we focused on the differences between the description of SR and ST in the texts regarding their Characteristics, Teleologies, and Associated Concepts. Although the differences between SR and ST are not as many as the points they have in common; with the current method, it is not possible to identify which of them are strictly shared or just overlap due to the conceptual shifting within the texts and authors using them as synonyms. Deeper content analyses of the texts are required to identify which constructs overlap and differentiate them from conceptual shifting. Generally speaking, SR was found to be a distinct concept for cognitive skills and cognitive processes associated with the scientific method. In the same way, the teleologies of SR are related to learning, to the development of hypothetico-deductive skills, and learning to perform steps of exploration related to the scientific method correctly. Concepts associated with SR were those related to experimentation, creation and development of scientific communities, and scientific achievement. ST was found to be rough-grained, focused on pattern detection, and development of models and scientific laws. ST was more often linked to the socio-historico-philosophical construct of science as a domain and used as a way to understand the world and to construct knowledge. ST was also found to be a concept used by non-psychological disciplines (e.g., nursery and game design) to refer to the use of the scientific method in their practices. Concepts associated with ST were also found to be socio-historico-philosophical, serving as a way of communication between scientific domains by means of metaphors.

2.8. Discussion

This study set out to clarify and discuss the characterisation and use of the concepts of SR and ST. The main research questions were (RQ1) Are SR and ST the same concept or different concepts? (RQ2) If SR and ST are not the same concept, what are their differences? (RQ3) How can we characterise and operationalise the constructs of SR and ST so that they can be more systematically used in research? This section discusses the findings made in the study integrating the discussion of the research questions with parallel findings made in the research.

2.8.1. Thinking and reasoning: An ontological problem

Some authors in Psychology, Philosophy, and Cognitive Sciences classify thinking and reasoning in different ways (e.g., Eysenck & Keane, 2003; Holyoak & Morrison, 2005). From a linguistic perspective thinking can be conceptualised in many ways; depending on the context, it can mean a state of belief, a plan, or judgement. Holyoak and Morrison (2005) define thinking as “systematic transformation of mental representations of knowledge to characterise actual or possible states of the world, often in service of goals.” (p. 2). This definition conceptualises thinking as a general macro-process which allows for detecting patterns and relationships, solve problems, making decisions, and developing conclusions. This process is related to high-end goals, critical, and creative thinking (Holyoak & Morrison, 2005, p. 776). Reasoning is defined as a specific process (Holyoak and Morrison, 2005), which is more specialised, deep, and narrow in its reach than thinking, but makes part of it. Being more specialised, reasoning is often classified into different types according to the type of task it is involved. Inductive reasoning (Eysenck & Keane, 2003; Johnson-Laird & Byrne, 1993), deductive reasoning (Garnham & Oakhill, 1994; Schoenbach, 1999), abductive reasoning (Eysenck & Keane, 2003; Johnson-Laird, 1983; Johnson-Laird & Byrne, 1993), hypothetico-deductive reasoning (Lawson, 2000), probabilistic reasoning (Lassiter & Goodman, 2015), and visuospatial reasoning (Aiello, 2002; Car & Frank, 1994) are amongst the types of reasoning typically explored in psychological research (Fischer, Hickey, Pellegrino, & Law, 1994; Johnson & Krems, 2001) and fostered in education (Aneta & Jerzy, 2013; Roberge, 1970).

These ontological perspectives on thinking and reasoning raise a question about the ontological nature of the SR and ST concepts and whether they are synonymous or if one is a process sub-type under the umbrella of the other, that is, SR as a subtype of ST. Based on the analyses conducted in the present study we argue that SR is not the same as ST: Therefore, they should not be used as synonyms. Theoretically speaking, thinking is a general ability used in everyday life to solve problems. Thinking is composed of other processes – i.e., reasoning and high-level cognitive processes (Gilbert & Burgess, 2008) – which interact in a complex way to allow for problem solving, decision making, and prediction. Moreover, some authors directly regard SR as a sub-process of Thinking, for instance, Coletta, Jeffrey and Steinert (2011), Etkina, Karelina and Villasenor (2007), and Popper (1959), specifically regard hypothetico-deductive reasoning as SR.

2.8.2. Scientific Reasoning is not Scientific Thinking

Accounting for the first research question, based on the analyses conducted , the ontological debate between reasoning and thinking, and that several authors explicitly describe their difference (Crowley et al., 2001; Dejonckheere et al., 2010; Dunbar, 2001; Dunbar & Klahr, 2012; Kisiel et al., 2012) we regard SR and ST as two different concepts that became entangled due to shared similar features (e.g., the use of the Scientific Method). Accounting for the second research question, SR is mainly a concept derived from Psychological and Educational research (see Figure 2) (Abdulkarim & Al Jadiri, 2012; Abdullah & Shariff, 2008; Inhelder & Piaget, 1958; Kuhn et al., 1992; Zimmerman, 2000) and is related to a series of cognitive processes or skills used to inquire and answer questions about the world and the nature of phenomena (e.g., create hypotheses, design experiments, collect evidence, relate variables, and create theories (Crowley et al., 2001; Dejonckheere et al., 2010; Dunbar, 2001; Dunbar & Klahr, 2012; Kisiel et al., 2012). In contrast, ST has a philosophical nature (Brigandt, 2010; Gower, 1997; Miller, 1983; Mormann, 2012; Popper, 1959, 1962) and is related to the content of science, scientific discovery, and scrutinising scientific discoveries (Crowley et al., 2001; Dejonckheere et al., 2010; Dunbar, 2001; Dunbar & Klahr, 2012; Kisiel et al., 2012). The descriptions of both concepts are in line with the ontological debate between reasoning and thinking, the first being more specialised and fine grained and the second being general and integrative. The separation between the concepts also reflects what we found in the network analyses, that is, although both concepts have characteristics in common, they also have unique characteristics; and even though the concepts are used as synonyms one-third of the time, they continue having differences.

2.8.3. Has Scientific Reasoning an implicit teleology?

In the present research we found different types of teleologies attributed to SR ranging from short-term particular goals such as performing better in school (Kuhn, 2009; Russ, Coffey, Hammer, & Hutchison, 2009) to long term broader goals such as fostering scientific discovery (Brown, Furtak, Timms, Nagashima, & Wilson, 2010; Schauble et al., 1991; Schunn & Anderson, 1999). Nevertheless, the fact that many contributions did not have an explicit teleology (particularly empirical research) raised a question about the value of SR. Does SR have an implicit teleological value? If that is the case, what is it?

The present study revealed that some researchers do not explicitly present the teleology of SR in their research, which raises the question if researchers on SR are departing from the intrinsic importance of SR? The ethical problem of not stating the teleology of SR, particularly when conducting empirical research, is that science should be in the position of avoiding dogmatism and view itself critically (Osborne, 1998). Academic community regards SR as a cornerstone cognitive resource which is intrinsically good. Historically speaking (Bourdieu, 1991) science has positioned itself in a place where observation, interpretation and knowledge acquisition are privileged cultural goods which possess socio-politico-economical influences. The Stanford Encyclopaedia of Philosophy (‘Scientific Method’, 2015) regards the scientific community as a community with high standards which often argues with other sciences for having neater and more linear methods of inquiry. Additionally, the “age of enlightenment” has played a key role in the contemporary occidental world, from power and ethics to knowledge and expertise (Osborne, 1998). From the perspective of ethics, it is required that researchers do not take things for granted, being sceptic with the phenomena they study, which includes their own ontological foundations. Despite the historical relevance of SR, it is important that community studying SR look at itself in a critical way instead of taking for granted its relevance, as it can be seen as a biased practice.

2.8.4. The necessity of characterising and operationalising variables

We found that many studies, particularly empirical research often do not characterise or operationalise their variables of study and/or conveniently base them on standardised tests. This proves to be confusing when evaluating the results of the research as it is not clear if the particular variables object of analysis has been reached during the study, or if general test results can be decomposed to specific constructs the researcher wants to study.

2.8.5. Scientific Reasoning, 'thinking like a scientist', and the Scientific Method

In the network analyses, we found a strong individual occurrence of the Teleology "Think like a scientist" and the associated concept of "Scientific Method" in relation to SR. This raised the question into what extent is SR a particular operationalization of the Scientific Method? If we take into account the definition of Scientific Method according to the Oxford Dictionary, it is a method consisting of (a) systematic observation, (b) measuring, (c) experimentation, (d) hypothesis formulation, (e) hypothesis testing, and (f) modification of hypotheses. If we compare it with the way Fischer et al. (2014) define SR, it includes (a) problem identification, (b) questioning, (c) hypothesis generation, (d) construction of artefacts –for measuring, (e) evidence generation, (f) evidence evaluation, (g) drawing conclusions, and (h) communication and scrutinising. The epistemic definition by Fischer et al. is a more detailed definition of Scientific Method but operationalised in a cognitive domain instead of a procedural knowledge approach. Additionally, we found the Teleology of "think like a scientist" linked to SR and the scientific method. To "think like a scientist" is not always clearly defined in the texts where the concept is mentioned, but generally speaking, the idea of thinking like a scientist is to use the Scientific Method (Kisiel et al., 2012). Based on the explored texts, we concluded that while the scientific method is regarded as a series of steps conducted for understanding a particular phenomenon, SR is regarded as the cognitive processes required to use the scientific method and is fostered by scientific enquiry processes.

2.9. Conclusions and directions for future research

Answering the first research question, SR was found not to be the same concept as ST; nevertheless, they are often confounded and even used as synonyms by researchers of the concepts. Answering the second research question, SR refers to a series of cognitive skills used to explore and understand phenomena in a systematic manner, whereas ST refers to the integration of scientific knowledge, discovery and socio-historical advance of sciences as a domain. We hope that our investigation on the usage of the two concepts in literature can inspire future work to more clearly separate the concepts.

Answering the third research question we propose that the concept of Scientific Reasoning, its characteristics and teleologies are grounded in hypothetico-deductive reasoning and can be operationalised in finer units, which allows for more controlled and systematic research in the future. In our view, the proposal by Fischer et al. (2014) is the most comprehensive, analytical and up-to-date definition and characterisation of the SR concept, which also includes argumentation processes as part of the scientific reasoning activity. This characterisation is an interdisciplinary effort (Learning Sciences, Psychology, Biology, Social work, Medicine, Computer Science, and Education) to understand SR and argumentation from different perspectives while including areas not very studied such as the role of emotions, the influence of social contexts, and the influence of communication media in the process of SR. The proposal by Fischer et al. is epistemologically directed identifying eight epistemic activities: (1) Problem identification, (2) questioning, (3) hypothesis generation, (4) construction of artefacts, (5) evidence generation, (6) evidence evaluation, (7) drawing conclusions, and (8) communication and scrutinising. Finally, the proposal by (F. Fischer et al., 2014) possesses clear teleologies or epistemic modes: (1) Theory building about social phenomena –centred on development of theories, (2) science-based reasoning and argumentation in practice –centred on development of tools and methods, and (3) artefactcentred Scientific Reasoning and Argumentation –circular approach in which the practice nourishes the theories, while the theoretical development nourishes the scientific practices.

For the concept of Scientific Thinking, we propose the use of the construct from Harre (2004) and Popper (1966) as the best characterisations. ST refers to a socio-historical process centred in the development of scientific theories and scientific content. Scientific development is not regarded as a collection of facts (truths) or evidence that allows science to improve but as a collection of falsifiable theories developed in historical contexts which influence their emergence and scope. The development of these theories is facilitated by the scientific community of the time, which also influences theories' span. Accordingly, ST is regarded as the collection of social and historical theories (provisional truths) allowing to trigger a shift in the scientific paradigms, the way the world is explained, the content of specific fields of science, and the theories and methods used in specific fields of science. Regarding the temporal nature of the characteristics of ST, it might also refer to the current advances in sciences by criticising and falsifying accepted theories.

Additionally, we suggest that when using the terms of reasoning and thinking authors refer to the constructs of reasoning and thinking proposed by Eysenck and Keane, (2003), and Holyoak and Morrison (2005). These constructs are clearly defined in terms of general and specific cognitive processes while accounting for their neuropsychological basis.

When studying SR, particularly in empirical studies, it is necessary to characterise the concept for correctly operationalising the variables. Even if the researcher is using a test, the factors of the test should be explained and related to the variables of study. In a similar way, empirical studies, particularly interventions, should describe the teleology of the interventions on skills for bringing a better and more critical perspective on the experimentation.

Although a separation of concepts was found using the network analysis, it is not clear to which extent these concepts truly differ and converge due to the conceptual shift of the texts. For future research we suggest to explore more in depth the concept of ST isolating it from the SR perspective by using the conceptual differences proposed by Crowley et al. (2001), Dejonckheere et al. (2010), Dunbar (2001), Dunbar and Klahr (2012), and Kisiel et al. (2012).

We suggest further studies using content analysis of texts for identifying which constructs overlap, into which extent, and the way this occur. The findings of the present research provide a ground for future research on the overlay of the constructs of SR and ST, by paving the way for steadier and finer studies.

There is a necessity to deepen in the relation between SR and the scientific method to understand better the nature of both and if there is a difference between both or if they are the same concept but operationalised in different domains.

Additionally, we suggest a replication of this research using a different and more extensive sampling method. The sampling method we used in the current research although fulfil its aim for a broad exploration of the two concepts in different domains of knowledge it is not exhaustive when accounting for specific fields of knowledge or deep database literature reviews; this opens the possibility for further exploration and construction of the concepts of SR and ST.

Finally, we suggest a genealogical study tracing the origins and ramifications of the concepts, as well as the theoretical references quoted by authors studying the concepts of SR and ST; this can shed light on why the concepts despite being ontologically different have been used as synonyms by the academic community.

Now that the concept of Scientific Reasoning has been critically assesses and characterised for a correct and standard operationalisation based on the proposal made by Fischer et al., (2014), it is possible to conduct the second phase of the study which inquires about the use of SR by MOBA players. This study is conducted in the following chapter.

Chapter 3

Scientists in the battleground: Scientific Reasoning and Strategy Making in Multiplayer Online Battle Arenas

Multiplayer Online Battle Arenas (MOBAs) gain popularity amongst the video gamer population as a consequence of their rise as an eSport (Hamari & Sjöblom, 2015). The current eSport tournaments move millions of people around the world (Popper, 2013), with prizes worth millions for the winners (Miller, 2010), and a whole community of professionally paid athletes who train every day to be the best at performing on these new type of sports (Tassi, 2012). Even some schools and colleges around the world are considering this new type of sports within their curricula, creating scenarios where students can participate in the old gym class as an eAthlete in a virtual environment (Maiberg, 2014; Tassi, 2015). From this, we can understand the salience of the phenomenon and how it spreads and influences the lives of millions of people (mostly teenagers and young adults) around the world. Due to the reach of these games, it is important to ask about their influence on the cognitive development of the players.

eSports comprise a broad spectrum of video game genres such as Real Time Strategy games, First Person Shooters, Fighting and Arcade Games, and Multiplayer Online Battle Arenas (MOBAs). As recent phenomena, eSports have not been widely explored by the academic community. Nevertheless, recent studies, particularly within the field of psychology, aim to understand how players play and which skills are used when playing (Bonny et al., 2016; Pöllänen, 2014; Wagner, 2006).

Traditionally, psychological studies in video games have focused on studying violence and addiction in games (Anderson, 2003; Anderson & Dill, 2000; Khang et al., 2013; King et al., 2013). Regarding positive human interaction with video games, there is a branch of Psychology and Game Studies research dedicated to investigating the impact of ‘serious’ and educational games on players’ behaviour and cognition (Connolly et al., 2012; McClarty et al., 2012). A smaller branch of psychology research investigates the positive behavioural and cognitive effects of playing commercial (off-the-shelf) games (Blumberg et al., 2013).

The present study, following the lines of Bonny et al. (2016) and Yang, Harrison, and Roberts (2014), explores which cognitive skills belonging to the area of Scientific Reasoning (SR) and Strategy Making (SM) are used and trained by MOBA players. MOBAs were chosen as a sub-genre of eSports for being highly competitive, fostering the exercise of high-order cognitive processes (Pereira, Wilwert, & Takase, 2016). Also, these games are popular and free to play which provides access to the game for a large population of players. SR and SM were chosen for three reasons: First, because these are cognitive skills acquired from a young age and developed and applied through life, either in an educated or a naïve way (Inhelder & Piaget, 1958; Sadler, 2004). Second, these abilities can be trained, exercised, and improved; they are not static. Third, because they are a cornerstone in problem solving (Holyoak & Morrison, 2005, p. 707), which is essential for overcoming obstacles in different environments, including the everyday life.

This chapter is structured as follows: First, the constructs of SR and SM are discussed, focusing on their relationship and their shared background in problem solving and executive function. Next, the research design, method and procedures are introduced. Subsequently, the results are presented divided into two sections, qualitative and quantitative analyses for both within-game and between-game differences, as well as a section with the triangulation of both data sources. Finally, the results and their implications for SR and SM are discussed, and conclusions are drawn.

3.1. Scientific Reasoning and its relationship with Strategy Making

SR, as described in Chapter 2, is a concept with different perspectives ranging from general problem solving (Abdullah & Shariff, 2008; Azmitia & Montgomery, 1993), to a specific set of abilities required to understand the nature of science, manipulate variables, design experiments, and interpret data (Koerber et al., 2005).

As described in Chapter 2, different epistemological lines have evolved concerning how to study SR. One line is derived from Popper (1959) and the Philosophy of Sciences; this line focuses on hypothesis systems (theories) built to inquire about the world. Theories are built from observable events and tested using experience. Popper defines SR as a deductive process where the scientist approaches a phenomenon with a hypothesis system, not to prove the system true, but to falsify it. Falsification is a quality conceding a hypothesis system to be tested for unfitness rather than fitness and allowing for constructing broader theories instead of proving correct the ones at hand.

The second line is derived from Inhelder and Piaget (1958) and focuses on developmental psychology. Inhelder and Piaget regard SR as a type of logical reasoning, based on theories of formal thought which are hypothetico-deductive in nature. Furthermore, they consider SR as the potential of reflecting about theories, build hypothetical models of reality, test interactions between variables, and evaluate evidence.

The present study uses the framework proposed by Fischer et al. (2014) which builds a comprehensive and interdisciplinary framework for SR based on these two lines. This framework concerns the present study in furtherance of the findings of the research presented in Chapter 2; it is the most updated definition of SR, developed by an interdisciplinary team of researchers accounting for different epistemic activities which allow defining the SR process in a fine-grained way, at the time that proposes clear teleologies for SR. The authors describe SR in terms of eight epistemic activities which are intertwined but not sequential: (1) Problem identification, (2) questioning, (3) hypothesis generation, (4) construction/redesign of artefacts, (5) evidence generation, (6) evidence evaluation, (7) drawing conclusions, and (8) communication and scrutiny of findings.

Theoretically, there are two different positions linking the processes of SR and Strategy Making (SM). The link between these processes is crucial for the current research as some of the cognitive activities linked to SR can be seen in-game in the shape of strategies. The first position linking SR and SM comes from studies in SR, which regards SM as a series of steps executed to falsify a hypothesis (Fischer et al., 2014; Inhelder & Piaget, 1958; Koerber et al., 2005). From the point of view of SR, SM is conceived as a strategy, a plan or a series of steps aimed at solving a scientific problem (Holyoak & Morrison, 2005; Klahr & Dunbar, 1988; Zimmerman, 2007).

The second position comes from SM, regarding SR as a basis for correctly assessing and deploying plans. Therefore, skills like control and manipulation of variables for testing hypotheses or falsifying information before making and deploying a plan are proper of SM (Martin & Olin, 1982; Wellman et al., 2005). Other authors regard SR to be closer to SM, linking it to the cognitive skills of induction, deduction, problem solving, probabilistic thinking, causation, and formulation of models (Etkina et al., 2007; Kagee, Allie, & Lesch, 2010; Popper, 1959).

Both SR and SM are intertwined at a deeper level; they are the product of a series of high-level cognitive processes which are the basis for problem solving (for instance, to engage in a task, make a plan for the future, switch tasks or inhibit a response (Gilbert & Burgess, 2008)). These processes are called executive function and are accountable for solving non-routine problems (problems which solution is not automated) which require analysis, attention, planning, separation of irrelevant stimuli, flexibility, information manipulation accounting for time, multitasking, prospective memory, monitoring actions, and control (Gilbert & Burgess, 2008; Rabbitt, 1997).

Thinking bottom up, SR and SM are interlinked at different cognitive levels. A healthy person possesses neurocognitive basis for executive function which is required to solve problems, particularly those that are new or which solution has not been automated. At the same time, problem solving is the basis for SR and SM.

SM is based on the definitions of problem solving and decision making (Kahnemann & Tversky, 1979; Newell, 1966; Newell & Simon, 1972; Tversky & Kahnemann, 1981) as a plan or a series of steps followed to achieve a goal. SM is a high cognitive process which ultimate goal is agency (the establishment of the actions required for the deployment of a strategy) (Jørgensen, 2003). SM has been characterised as a metacognitive and metarepresentational skill by which a person can understand the diverse elements and actors within a problem and, based on past and present information, create a model of the problem allowing for its solution. The strategist needs to make systematic observations and project different probable outcomes based on the interaction of the elements composing the problem so a course of action can be drawn (Hong & Liu, 2003; Siegler, 1999).

SM is also related to Theory of Mind (Colman, 2003), particularly when the agents are humans. In these cases, the course of action is not only determined by understanding the different pieces of information corresponding to the problem, but also the history, intentions, and probabilities of action of the agents interacting with and within the problem.

SM, as stated by Miller (1989) has multiple process and dimensions, but the three-dimensional models are the ones most commonly cited. The first dimension of SM is information processing, which implies active gathering and use of internal and external information. Miller proposes a division of this dimension into four sub-dimensions: Analysis, planning, strategy creation, and environment scanning. The second dimension is interaction, implying decision-making product of interaction with the environment; this dimension is divided into two sub-dimensions: Consensus building/individual decision, and bargaining. The third dimension is assertiveness, which concerns the level of risk of a decision and whether this is reactive or proactive; this dimension is divided into two sub-dimensions: Proactiveness and risk taking.

Following these relationships, the epistemic activities of SR proposed by Fischer et al. (2014) can be correlated with the three dimensions of SM proposed by Miller (1989). Accordingly, information processing corresponds to the epistemic activities of observation and identification of a problem (e.g., field assessing in military strategies), hypothesising about different element interactions and outcome scenarios (e.g., the next possible move of the black chess piece and the balance of the board afterwards), construction of artefacts (e.g., the game *Don't Starve* (Klei Entertainment, 2013)), information gathering (e.g., the Wason card selection test). The interaction dimension corresponds to the epistemic activities of evidence generation and evaluation (e.g., the game 'Mastermind'), drawing conclusions and communicating either the elements found or the decision made (e.g., intelligence research in military strategies). The assertiveness dimension is the place where SR and SM differ the most, as it implies the deployment of a strategy or a process under uncertainty and risk. SR does rarely possess that faculty as it is a spiral process going over itself due to its ontological characteristics of scrutiny and reflection.

3.1.1. Problem types and Problem Solving

Historically, there are two main lines in assessing SM in problem solving: Algorithmic and heuristic problem-solving (Newell, 1966; Newell & Simon, 1972). Algorithmic problem-solving consists on conducting a series of steps starting from a defined situation, and considering affordances and restrictions to achieve a goal (e.g., going up a staircase (Cortázar, 1969)). Heuristic problem solving consists of detecting patterns within a problem, which allows solving the problem skipping some steps from the algorithmic solution (e.g., going up a staircase jumping the steps in groups of two).

Problems can be categorised as well-structured and ill-structured problem-types (Jonassen, 1997, 2000; Montealegre, 2007). Well-structured problems are organised, with a clear starting point and well-defined affordances, restrictions, and goals. These problems can be tackled algorithmically; and with expertise, heuristics can be created to solve these problems more efficiently – e.g., going up a staircase (Cortázar, 1969). Ill-structured problems, do not present clear starting points, affordances, constraints, procedural steps, or goals. Typically the 'solution' to an ill-structured problem consists on achieving equilibrium between benefits and drawbacks under certain conditions.

Dado (2015) relates problem solving with situation awareness as a precursor and the main piece in problem solving, particularly for ill-structured problems and video games. Situation awareness can be defined as the set of cognitive skills required to work in or control a dynamic environment (Durso & Gronlund, 1999). It entails information acquisition from the environment and its integration with new and previous information to formulate a correct representation of a changing context, allowing for predicting its future states (Endsley, 1990). Following Dado's line of awareness in video gaming, 'map awareness' and 'game awareness' appear as specific instances of situation awareness where a person (in this case the player of a game) acquires information from the problem space (map) or the interaction between elements of the environment (game); then creates a representation of the game context for identifying patterns, which allows for heuristic problem solving or mapping the branches of future events in a probabilistic manner.

3.1.2. Video Games, Problem Solving, Strategy Making, and Scientific Reasoning

The process of problem solving has been a classic field of study in Psychology (Newell, 1966; Newell & Simon, 1972) and decision making in economics (Kahnemann & Tversky, 1979; Tversky & Kahnemann, 1981). Games, in general, possess the structure of a problem, either well-structured or ill-structured. Besides, games propose a structure of micro and macro problems that need to be solved to progress, complete, and succeed on it (Jørgensen, 2003). Take for example the case of chess, the game starts in a status-quo where the black and the white pieces are on a balance, the main goal is to break the balance by surpassing the opponent team (macro-problem), meanwhile, small problems need to be overcome so the player can advance towards the bigger goal.

Video games are not an exception to this structure as they behave in the same way. Some video games are well-structured problems (e.g., Sokoban (Thinking Rabbit, 1982)) depicting a series of actions (or mechanics, such as move and push) to follow to beat a level (micro-problem), and beating a series of levels to beat the game (macro-level). Other video games are ill-structured, and the player can overcome a situation in different ways, which purposes are not to ‘win’ the game but to achieve a balance allowing them to progress (e.g., Dragon Age (Bioware, 2009)).

MOBAs behave as ill-structured problems. Although the game possesses a set of rules and algorithms, as well as certain status-quo start (the status-quo is not total as different players chose characters with different skills) and a defined goal, the action possibilities and problems that can emerge during a match are uncountable. To solve the challenges posed by the game, players use a combination of SM and SR. The SM in these games consists of gathering as much information as possible and, based on the affordances and restrictions of the game (in other words, in the game rules and mechanics), make the decision that favours the player or the team the most. SM in MOBAs can be seen regarding the dimensions proposed by Miller (1989); thus, information seeking from the team and the opponent, as well as mapping of the enemy position and affordances correspond to the Processing Dimension. Making and contrasting hypotheses about the enemy behaviour, and planning for an attack or counter-attack correspond to the Interaction Dimension. Communication of possibilities, outcomes, and mistakes correspond to the Assertiveness Dimension.

In the same way, SR can be seen in MOBAS when players gather information and make hypotheses about the enemy behaviour and possible outcomes (problem identification and questioning). They do so either based on known patterns or a tree of possibilities (Newell & Simon, 1972) and test their hypotheses using special constructs given by the game or player communication (hypothesis generation). Additionally, MOBA players can set beacons to locate the enemy (construction of artefacts), or probe the terrain and the enemy gathering information to contrast their hypotheses and narrow the outcome possibilities (evidence generation and evidence evaluation). Players communicate their findings via chat or pings (communication) and act based on this information, also evaluating the outcome of their actions once they have come to an end (drawing conclusions).

A way to approach the study of SR in video games is studying the difference between player's cognitive skills according to their level of expertise in a particular game (Bavelier, Green, Pouget, & Schrater, 2012). To understand this comparison, we deepen the understanding of expertise in the following section.

3.1.3. The role of expertise in Scientific Reasoning, Strategy Making, and games

As pointed by Schrader and McCreery (2008), there is an extensive research in different fields regarding the nature and characteristics of expertise. One of the most used frameworks to characterise expertise is that of Chi, Glaser, and Farr (1988) consisting of seven characteristics differentiating experts from novices: Experts best mainly in their particular domain, distinguish meaningful patterns within their domain, are fast and efficient solving problems within their domain, have greater short-term and long-term memory, represent problems at a deeper level when compared to novices, spend a long time analysing problems qualitatively, and have strong self-monitoring skills.

Studies in expertise have been conducted in different domains, from medicine (e.g., Lawson & Daniel, 2011) to sports (e.g., Furley & Wood, 2016). Nevertheless, as the focus of this research is on SR and video games, we are centring on expertise in these domains. Schunn and Anderson (1999) made a study regarding expertise in *memory* comparing three levels of expertise according to the time spend researching on memory. These studies aimed to explore the generality/specificity of expertise of SR, finding that some of the variables measured were only attainable by experts on the field; thus, accounting for expertise as highly domain-specific. Besides, Schunn and Anderson also found that domain-specific expertise is not the only factor important in the development of expertise, but domain-general expertise also influences the way in which experts behave in domain-specific tasks.

Schrader and McCreery (2008) focus the expertise research in Massive Multiplayer Online Games, shedding light over the characteristics of expertise in games, which differ from expertise in other fields of research. One of the points explained by Schrader and McCreery is that the traditional model of expertise based on a mentorship model does not fit video games; instead, it obeys to a model of dynamic interaction between players at different levels. Such levels integrate different aspects of the video game environment, like use of peripherals (i.e., keyboard, mouse, game controller, and joystick), game mechanics, game content, specific procedures, social interactions with other humans (e.g., the creation of guilds), game plot, and the use of external tools (e.g., TeamSpeak) to communicate while playing. Due to the complex and dynamic nature of games, expertise is developed at many levels and in many ways, such practice, mentoring, vicarious ‘practice’ though video stream, and the creation of communities of practice and knowledge (Martin et al., 2011; Steinkuehler, 2004, 2006). Furthermore, Schrader and McCreery point that players improve their knowledge of the game and their performance in the game according to the level of expertise.

Hong and Liu (2003) propose a tree step model to classify acts of proficiency on expertise in game players, trial-and-error- thinking mode (run across a problem and solve it), heuristic-thinking mode (use of mental models to solve the problem), and analogical thinking mode (use mental models and previous experience to create mental scenarios). As an example, when players play a video game for the first time, they try the mechanics of the game for understanding how they can interact with the environment. When players have a grasp on the mechanics, they create mental models and schemata that can implement in other parts of the game (i.e., advanced levels). Later, when the player plays another (similar) video game, they try to use these mechanics and schemata with the new game and adapt it to the new situation. This process is very similar to the processes of assimilation and accommodation described by Piaget (1964).

3.2. Research aims

The present study explores four Research Questions concerning SR in MOBAs: (RQ4) How MOBA players use SR and SM when playing, (RQ5) whether differences in the level of expertise of players delineate patterns SR and SM on their play, (RQ6) whether there is a difference in both SR and SM according to the game played, and (RQ7) the relationship between SR and SM when playing MOBAs.

3.3. Design

The present paper reports on an exploratory study involving expert MOBA players of the games League of Legends (LoL) (Riot Games, 2009) and Defence of the Ancients 2 (DotA2) (Valve Corporation, 2013). The study uses a convergent mixed methods design (Creswell, 2015) by triangulating (a) qualitative data regarding the use of information in the game, hypotheses generation, hypotheses testing, and strategy making, and (b) quantitative data obtained with eye-tracking technology using fixations as an indicator of attention and information processing, and saccades as information search (Goldberg, Stimson, Lewenstein, Scott, & Wichansky, 2002; Goldberg & Kotval, 1999; Just & Carpenter, 1976; Poole, Ball, & Phillips, 2004; Poole & Ball, 2005). Additionally, it makes use of an exploratory design (Stebbins, 2001) which allows for understanding and identifying behaviour and indicators of SR and SM in MOBAs.

3.4. Setting

Gameplay sessions took place in a simulated-natural environment at the game research laboratory of the Catholic University of Applied Sciences Munich (see the concept of 'Living Room Lab' in Lieberoth & Roepstorff, 2015), meaning that participants were instructed to play in the same way as they typically play at home; therefore, we allowed the use of playing aids (e.g., professional gaming mice or keyboards) and the modification of the game play settings and Head-Up Display (HUD). Modifications to the main game (*mods* or *cheats*) were not allowed. The major visible difference to a simulated natural (home) environment was the use of eye-tracking devices.

Participants were instructed to play in pairs with a fixed playmate of the same or closest level of expertise. Furthermore, players were asked to talk with their partner and share any strategies or thoughts just as if they were having a talk with their playmates during a normal game.

3.5. Participants

A total of 28 players participated in the study. Eighteen players between the ages of 18 and 26 years old (7 female, 11 male), participated as LoL players. Ten players between the ages of 20 and 26 years old, all of them male, participated as DotA2 players.

Participants were selected using convenience sampling. The inclusion criteria for the study were the following: (a) Speak fluent English and be willing to participate in the in-game talk and interviews in English language, (b) have perfect or corrected to perfect vision (so that vision problems do not interfere with the eye tracking device), and (c) agree to the terms of the informed consent and the specified payment. Also, for LoL players, it was required to be ranked in the LoL Ranking system (‘League system’, 2016). For DotA2 players, it was required to be ranked in the DotA2 Elo point system (Elo, 2008; ‘Matchmaking’, 2016).

The participants were divided into different categories according to the respective ranking systems for their pairing. For LoL, we used Riot Games’ ranking system (the system used at the time of the data collection, November 2014 to April 2015). The system is based on Tiers and Divisions. There are six Tiers: Bronze, Silver, Gold, Platinum, Diamond, and Challenger. Each of the first five Tiers is divided into five Divisions (I, II, III, IV, and V) forming a continuum of 25 Division ending in the final Challenger Tier. We excluded the

Challenger Tier for being reserved for professional players, whose participation was difficult to get for the study. Based on this system the players were re-grouped into eight categories (A thru H; see Table 12) to simplify the identification of suitable pairs. Additionally, we selected this division for securing regular intervals allowing for a balanced group evaluation (not fine-grained and not rough-grained). The distribution of participants for LoL is presented in Table 12.

Table 12

Categories created for League of Legends analyses based on Tier and Division and distribution of participants according to categories

Tier	Division	Category	Number of participants	Percentage of participants
Bronze	I-IV	A	0	0.00%
	V	B	3	16.67%
Silver	I-II			
	III-V	C	2	11.11%
Gold	I-III	D	1	5.56%
	IV-V	E	8	44.44%
Platinum	I			
	II-IV	F	0	0.00%
	V	G	1	5.56%
Diamond	I-II			
	III-V	H	3	16.67%
Total			18	100.00%

The DotA2 ranking system is based on the Elo point system (Elo, 2008). As there are no fixed categories in the Elo system, categories were created based on the points of the participants. Six categories were created (I thru N; see Table 13) to simplify the identification of suitable pairs. Categories were created starting from 3000 points (lower point total for participants) up to 5001 or more points (higher point total for participants) and in regular intervals of 500 points. The distribution of participants for DotA2 is presented in Table 13.

Table 13

Categories and distribution of DotA2 players based on the Elo point system

Elo points	Category	Number of participants	Percentage of participants
3000<=	I	1	10.00%
3001-3500	J	3	30.00%
3501-4000	K	2	20.00%
4001-4500	L	3	30.00%
4501-5000	M	1	10.00%
5001>=	N	0	0.00%
Total		10	100,00%

3.6. Apparatus

This section is divided into four parts, and it describes the apparatus used: The games (similarities and differences), the computers, the software used for data capture, the computers, and the eye-tracking devices. Concepts and jargon associated with the game will be emphasised in italics; a comprehensive description can be found in Appendix F.

3.6.1. The games, similarities and differences

The games used in the research (LoL and DotA2) were chosen for their structural similarities and popularity. Following we present a detailed description of similarities and differences between games.

3.6.1.1. Similarities

Both games consist of a system of *heroes* with different abilities and strengths. According to their characteristics, different heroes can be used to perform a *role* in the game (e.g., *marksman*, *support*, *tank*, or *jungler*). Players of the games choose a hero based on their individual, team strategy, or role preferences at the start of the game. The game takes place on an isometric map, and the goal is to defend the team base while trying to destroy the opponents' base (*Nexus*-LoL/*Ancients*-DotA2). To achieve this, the players get help from *minions/creeps* and must work together as a team to protect each other while destroying enemy structures (*Turrets*/*Towers*). Later on in the game, they can unlock *super-minions/super-creeps* by destroying one of the enemies' final protective structures (*Inhibitors*/*Barracks*). Players gain *levels of experience* and *gold* depending on their performance on the field. With the *experience* gained, they can unlock and improve *skills* used for attacking and defending. With the *gold*, they can buy items that allow them to perform different attacks or increase some of their base *stats* (see Appendix F).

The game purpose is similar to a 'capture the flag' game, but instead of capturing a flag and returning to the base, the team must open its way amongst offensive structures and tackle down the opposing team to destroy the main structure (*Nexus*/*Ancient*) at their enemy's base. The team winning is the first one on destroying the *Nexus/Ancient* at the opponent's base.

A *ranked game* consists of two teams formed by five people each, fighting for the control of the terrain (map) over a three-lane isometric map. The map is divided according to specific area locations; it is counterbalanced, so no team has an advantage over the other.

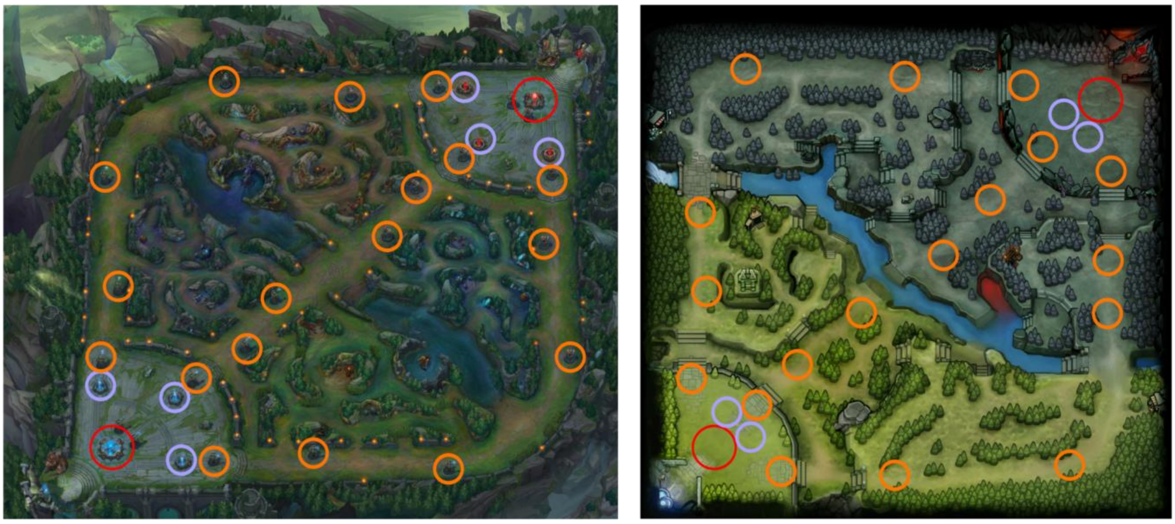
The HUD of both games is similar, as both depict (a) the game battlefield with unknown areas covered by *Fog of War* (enemies or threats are not depicted, but only a rough representation of the geography of the map, see Figure 8 for an example of the players' field of view), (b) a *minimap* (equally affected by the Fog of War) containing the overview of the game such as the location of allies and enemies (see Figure 7 for an example of the game minimap), (c) a list of purchased *items*, as well as the keyboard shortcut to use each item (see the yellow squares in Figure 2), (d) the *skills* of the character as well as the keyboard shortcuts to use each skill, the current level of the *abilities*, the amount of health (*life*) the character has, and the *magic* or *power* points to spend (see the blue squares in figure 2), (e) the current *stats* of the player and the rough status of the teammates, and (f) an additional window that can be opened to view the amount of *gold*, *kills*, and *items* of the teammates and enemies. Additionally, special *skills* or *items* can enable players to see through the *Fog of War* both on the battlefield and the *minimap*. Allies' positioning also gives vision to the player in the minimap, as the players have *shared team vision*.

Communication-wise the games have two main communication systems: Written chat and *pings*. The chat consists of a regular text chat panel where the player can write and read messages for their team. *Pings* are visual awareness indicators displayed on the game map and minimap and depict information such as "warning", "go back", "missing enemy", and "help".

Both games possess different ways to provide temporary ability boosts (*buffs*, see Appendix F) to players of their team by either contesting a place or killing *neutral enemies* (*neutral creeps*, see Appendix F) on the battlefield. *Neutral creeps* have specific locations, appear at different time intervals (*spawn times*, see Appendix F), and possess different *levels of difficulty*. The higher the difficulty of the *neutral creep*, the bigger the *gold bounty* and the *buff* provided by it.

Structural similarities between the game maps are presented in Figure 7. Structural similarities within the HUD are presented in Figure 8

Figure 7. League of Legends (left) and Defence of the Ancients 2 (right) maps depicting the location of game structures



Note. The red circles indicate the Nexus/Ancients; the light-blue circles indicate the Inhibitors/Barracks; the yellow circles indicate the Turrets/Towers.

Figure 8. League of Legends and Defence of the Ancients 2 Game-Play Screens Highlighting the game HUD

Note: LoL play-screen to the left, DotA2 play-screen to the right. The red square indicates the minimap; the blue square indicates the ability bar; the yellow square indicates the inventory.

3.6.1.2. Differences

Although the games are structurally similar, there are small differences influencing how players connect with and make use of the game, which at the same time influences how they engage in SR and SM practices within the game. Regarding *game mechanics*, there are four main differences between LoL and DotA2. First, while LoL has one *store* at each base where players can buy their items, DotA2 has two stores. One of the stores is close to each base where players can buy basic items for building up their hero, and a "hidden" store, on each side of the map, where more powerful items can be bought. As these hidden stores are on neutral terrain, they can be contested and controlled by both factions; thus, controlling them not only offers more powerful items but simultaneously prevents the other team to further develop their heroes. The possibility that resources can be withheld from players influences SR and SM to the extent it forces players to change their strategy either to take advantage over the other team or to defend their resources, so their growth is not slowed down.

Another difference in mechanics is that LoL has set *lanes*, and only specific abilities allow players to go through walls from one lane to the next. In DotA2 the boundaries between lanes are more diffuse as there are items and skills that allow a player to destroy parts of the barrier between lanes, thus creating paths that alter the structure of the map. Moreover, DotA2 is built in such a way that the battlefield has depth; accordingly, some places can only be accessed by *flying* characters or by *teleporting* to the specific place. Having a changing environment affects SR and SM to the extent that it forces players to seek for more information about the changes of the environment, keep track of them, and adapt new strategies of attack and defence accounting for the restructuration of the map.

The third difference in mechanics is that LoL has heroes with set *stats* (see Appendix F) and abilities for a specific role, whereas the roles in DotA2 are more flexible as most of the abilities of the *heroes* are derived from items, and their stats are not defining. Thus, a *hero* in DotA2 can switch from attack to defence easier than a *hero* in LoL. This flexibility difference allows players of LoL to sustain a certain individual and group strategy for most of the game, while DotA2 players are confronted with a more dynamic environment with more variables they need to adjust.

The final difference in the mechanics is that when a DotA2 *hero* dies in the game, the opponent *killing* the player's hero gets *points* for killing and steals their *gold*. In contrast, a *killed* hero in LoL does not lose gold. This mechanic in DotA2 acts as a negative punishment, fostering a more competitive behaviour and encouraging the use of SR and SM by players. As a note, a hero has infinite lives and can re-appear (respawn) at the Nexus/Ancient a certain time after killed.

The HUD is very similar in both games, except that in LoL the HUD elements are displayed in opposite order as compared to DotA2. Nevertheless, the game options of both games allow for changing the size and position of many elements of the HUD, so it is highly customisable.

As for the communication, DotA2 has an integrated voice chat, so the players can communicate with their team without stopping their actions in the game and shift their attention to writing in the chat. The ease of communication inside the game allows players to focus their attention on the SR and SM processes involved in the gameplay instead of forcing the player to shift attention for writing in the chat. LoL does not have an integrated voice chat, so players can only communicate via voice using external programs; this makes that LoL players can only communicate via voice with other players they already know, while DotA2 players can communicate via voice even with players they do not know.

3.6.2. Computers

Four computers were used during the play sessions, two gaming laptops for the players and two eye-tracker dedicated computers for the researcher. The gaming computers used

Microsoft Windows 7 as their operating system and ran the games and software used to record screen captures.

The eye-tracker dedicated computers used Linux Ubuntu as their operating system and were used to run the ‘PupilCapture’ software during the play sessions. They were also used to pre-process the data and render the eye-tracking videos using the ‘PupilPlayer’ after the play sessions.

3.6.3. Eye-tracker

Two ‘Pupil Pro’ head-mounted eye-trackers were used simultaneously; one for each participant of the dyads playing. The eye-trackers were designed by Pupil Labs (Pupil-labs, 2014) with the following specifications: Video-based combined with pupil/corneal reflection, point of regard measurement, eye camera: Maximum Resolution - 640x480 @ 30fps (60 Hz), world camera: Maximum Resolution - 1920x1080 @ 30 fps (60 Hz), infrared camera with IR Filter. The cameras of the eye-tracker record the pupil and corneal reflection of the eyes of the participants and triangulate it with the field camera that records the game screen.

Each gaming computer was paired with an eye-tracking computer. Both computers were placed beside each other to calibrate the eye-trackers more easily and store the data without disturbing the player. Each participant was located in a different section of the research laboratory, wore the eye-tracker as well as a headset to hear the game sound without interfering with their communication acts. The headset was fitted with a microphone, and their communication was recorded by FRAPS (Fraps & Beepa, 2013). In the case of DotA2, the microphone also recorded the in-game voice chat.

3.6.4. Software

Two pieces of software were used for data capture. FRAPS was used to record the game screen/gameplay and talk of the players. The software PupilCapture v4.5 (Pupil-labs, 2014) was used to record the eye-tracking data; it records both the corneal reflection of the player and the view from the perspective of the player towards the screen. An additional digital voice recording device was used to record the voices of the players so communication between them could be analysed. The software PupilPlayer v4.3 (Pupil-labs, 2014) was used to read and pre-process the data obtained with the eye-tracker. This software allows for the creation of Areas of Interest (AoIs) based on QR code markers, as well as post-processing of eye-tracking calibration, merge and export the eye-tracking data and the gameplay videos, and export pre-processed mathematical eye-tracking data.

3.6.5. Interviews

Semi-structured interviews were conducted with the participants the first day of participation. The semi-structured interview was designed to inquire the participants about their strategy, the use of elements of the game and the HUD, as well as the interaction with other players. The questions of the semi-structured interview were oriented by the subjects we were inquiring for in the research, as well as by elements which could be paired with the eye-tracking data. The semi-structured interview also served the purpose of building rapport with the participants (facilitating the structuration of the setting) and knowing how players understand the game, how they connect with it, how they build their strategies.

3.7. Variable operationalisation

For the present research, we defined SR based on Fischer et al. (2014) epistemic activities. Also, we defined SM based on Miller (1989) three-dimensional model: Information (map awareness, game awareness, and use of the minimap), interaction (evidence generation and evidence evaluation), and assertiveness (which strategies do player have and how they implement them). We operationalised the variables for both the qualitative and quantitative research based on these two definitions.

For the qualitative analyses, we conducted a semi-structured interview asking for the main cues related to the process of hypothesis generation of the players, the use of the minimap, the process of strategy making, and the communication process. These cues were lately subdivided in an inductive way accounting for fine-grained aspects related to SR and SM described in the previous paragraph.

Qualitative data was analysed looking for general patterns of action inside the games as well as differences between game players of both LoL and DotA2. Being an exploratory research the main goal of the qualitative research is to understand patterns of actions and decisions made by the players, allowing us to understand the processes of SR and SM of gamers. Additionally, the qualitative methodology allowed differentiating some of the quantitative measures which could not be quantitatively separated. As an example Fixations per Second are associated with three activities (problem identification, hypothesis generation, and evidence evaluation); to discern which activity is more prevalent a qualitative analysis is necessary.

Quantitative analyses were based on eye-tracking data. Theoretical research on eye-tracking (Just & Carpenter 1976; Goldberg & Kotval, 1999; Poole & Ball, 2005; Duchowski, 2007) indicates that eye fixations (Fixations per Second, as measured in the present study) are related to time spent processing and integrating old and new information, thus accounting for information processing. High amplitude saccades (Saccades per Second, as measured in the present study) are related to information search (also called Map Awareness and Game Awareness in the present study). The average of fixation duration (Mean Duration of Fixations) is related to the time spent recognising and processing new information (called Information Process in the present study).

Fixations per Second (FpS) are related to problem identification, hypothesis generation, and evidence evaluation. Saccades per Second (SpS) are related to problem identification, information search, map awareness, game awareness, and evidence generation. The Mean Duration of Fixations (MDF) is directly related to executive function regarding the time required to identify new information.

The quantitative data of the study aims to support and contrast the findings of the qualitative research while accounting for differences in the way players reason according to their level of expertise. The aim is not to compare differences between expertise levels of players, but using the expertise levels to detect and understand patterns of SR and SM on players.

3.8. Procedure

Participants were acquired via social network groups of students and gamers within the Bavaria region in Germany, as well as an open call using posters in public places. Interested participants contacted the research team and were asked to answer a questionnaire to assess whether they met the inclusion criteria and to correctly pair them for the play sessions according to game and expertise.

Each playing session consisted of two players situated in the same room playing cooperatively. The semi-structured interviews were conducted before every play session. Afterwards, the software for data capture and the games were checked for new updates. Following this procedure, special hardware was installed if the players brought any. The games were set up using the participants' accounts so that they had access to their personal profiles. Players were given time to prepare the settings of the HUD for them to feel comfortable. Although the HUD differences do not directly impact the gameplay, if a player is confronted with a different HUD than the one they are used to, they might have problems finding information on the screen or navigating the map (thus, interfering with SR and SM of the players). To prevent interference, we encouraged that they arranged the HUD as they were used to. The eye-trackers were calibrated using the automated screen marker calibration feature which consisted of participants following with their sight a marker moving through nine points on the screen. Afterwards, the procedure was repeated up to two more times to ensure an accurate calibration. Additionally, the calibration procedure was performed every time a new game started to ensure accuracy. Players regarded the eye-tracker as nonintrusive when they were playing and reported that the setting let them feel like an internet café were game tournaments take place.

Two play sessions were recorded over the course of two days for each pair. The pairs played multiple games during each play session. LoL players played three ranked games. Ranked games take place in a three-lane map consisting of two teams of humans against each other (see Appendix F). DotA2 players played five ranked games. Each player (irrespective of the game) played an average of 5 hours in the two days of playing.

3.8.1. Measures

Descriptive statistics scores were processed using the package "Psych" v1.4.3 for R. Table 14 describes the characteristics of the participants and the initial number of observations (games played) according to the game played.

Table 14

Characteristics of the participants according to the game played

Game	Number of participants	Initial number of observations	Minimum age	Maximum age	Won games	Lost games
LoL Players	18	54	18	26	25	29
DotA2 Players	10	50	20	26	24	26

3.8.2. Interviews

Qualitative content analysis was used to analyse the interviews. The semi-structured interviews were recorded using a digital voice recorder. The recordings of the semi-structured interviews were transcribed using a Pure Verbatim format, thus accounting for the literal expressions of the players as well as other utterances and onomatopoeic expressions. The qualitative content analysis was a mixed type, combining inductive category formation and nominal deductive category assignment (Mayring, 2014). The interviews were first separated according to the population interviewed (LoL participants and DotA2 participants) and then separated according to the questions. In the first step of the analysis, semantic units of analyses were defined as statements of different lengths and which refer to the core elements of the research (SR and SM). The statements were acquired from the transcribed interviews and were assigned by the researcher to one of four nominal, exclusive, deductive categories based on the research aims: Hypotheses, Minimap, Strategy Making, and Strategy Communication. The creation and assignment of units of analysis were supervised using the audit system proposed by Akkerman et al., (2008). In the second step of the analysis, statements were used to create non-exclusive inductive categories derived from the core elements of the deductive categories as follows: For hypotheses making: Hypothesis generation, Hypothesis topic, Hypothesis occurrence, Hypothesis testing, Interpretation of results, Hypothesis backing, Hypothesis goal, Information sharing, Intervening factors, Integration of hypotheses into strategy, Map awareness, Game awareness. For minimap use: Seeking general information, Seeking specific information, Navigating and exploration. For strategy making: General Precursory Strategy, Specific Precursory Strategy, Flexible Precursory Strategy, In-Vivo Strategy, Absence of a strategy. For Strategy communication: Written chat, Voice chat, Pings, No communication. Table 15 provides an overview of the deductive and inductive categories formation and their description.

As a note, examples taken from the interviews are annotated inside quotation marks followed by the source in parenthesis. The source of each participant is described by a four characters code, the first two characters are numbers representing the participant number, and the second two characters indicating if the participant was a LoL player (LO) or a DotA2 player (DO).

Table 15

Description and examples of the categories used to analyse the interviews

Deductive category			
Hypotheses	Assigned to the question "Do you make hypotheses when you play?"		
	Inductive category	Description	Example
	Hypothesis generation	Statements as to whether the player makes hypotheses or not while playing	"Yes, ehm, not always consciously, but I try to do it more and more as I am progressing in the game" (18LO)
	Hypothesis topic	Statements about the topic of the hypotheses the player makes	"Predicting what enemy, where enemy could be, and what they should do" (30DO)
	Hypothesis occurrence	Event or time triggering the hypothesis generation on the player	"If all heroes are missing" (34DO)
	Hypothesis testing	The player does something to test their hypothesis or examine the results of a probe they sent to gather information	"Well if we get killed or something, I can see that it was the wrong decision" (05LO)
	Interpretation of results	The player interprets the results of a hypothesis testing or a certain information pattern on the game	"'Not seeing people' that is danger, 'cause they could be anywhere." (21LO)
	Hypothesis backing	The player makes claims to back their hypotheses.	"Because then we are lesser than they are, so we can't defend the tower" (36DO)

(continued)

Table 15 (continued)

Hypotheses	Inductive Category	Description	Example
	Hypothesis goal	Statement about the aim or goal of making hypotheses	"You are not losing the fight, but if you are predicting them, and you are predicting them right, you are 100% going to win" (10LO)
	Information sharing	The player share their hypothesis or results of probing with the rest of the team	"I try to warn them, [sniff] [tongue click] yeah" (11LO)
	Intervening factors	If there are conditional factors or a chain of events that triggers or confirms a hypothesis	"If I, if I see my, my mid-laner, ehm, is going back through the Fog of war, I try to predict if he's going to gank, if it's bot' or top lane" (11LO)
	Integration of hypotheses into strategy	Players integrate their hypotheses, tests and probes into their strategy	"I think, hypotheses, they have this guy on wait, he will make some items, and to counter those items, I should do this kind of items." (30DO)
	Map awareness	The player implicitly or explicitly refers to map awareness when describing their hypothesis or strategy	"To see what vision we have, or very often erh, to see if I can go in, or not, if I can initiate." (26DO)
	Game awareness	The player implicitly or explicitly refers to game awareness when describing their hypothesis or strategy	"To have an overview of the situation-- of the game situation, to see whether-- how my teammates are doing on their lanes, or to see if there is some important objective where I could actually do something useful or where my help is needed!" (03LO)

(continued)

Table 15 (continued)

Minimap	Assigned to the questions “How do you use the minimap?”		
	Inductive category	Description	Example
	Seeking general information	The player uses the minimap to gather new (general) information	“To see that my team has people who are low at health -- sometimes just to see what is happening” (06LO)
	Seeking specific information	The player looks at the minimap looking for a specific piece of information	“Ehm, I use the minimap. I have, ehm, every hero has its own little picture on my minimap, so I'm not only seeing which hero is, is evil to me, but I also seeing which hero/here(?) it is. So I'm not looking at the minimap, I have red and green points, I have little pictures. So it's really, really good to know which heroes are coming up, because you have to instantly know what they can do, why they're here. Because if you see some heroes on the lane, you wouldn't expect them to be, it's always suspicious and you have to be careful.” (36DO)
	Navigating and exploration	The player uses the minimap to move or look around	“Uhm, yeah [giggle], I look at it, and I just, uhm, navigate by the mouse” (32DO)

(continued)

Table 15 (continued)

Strategy making	Assigned to the question "Do you have a strategy when you play the game?"		
	Inductive category	Description	Example
	General Precursory Strategy	The player has a very basic strategy before starting a match	"Because when heroes are missing, ehm, you have to care by your, erh, all heroes are ehm, erh, erh, erh, example are down, you can push up, because nobody can gank you. Erh, so it's very important to watch the minimap, but you also, ehm, ehm, much different, because erh, you ehm, need the last hit, and concentrate on last hitting, and watching the minimap. So if you're too concentrated on last-hitting, you can easily ganked erh, from heroes out of nowhere, because you don't watch the minimap. So that's the main, main strategy." (34DO)
	Specific Precursory Strategy	The player has a specific strategy of how to counter and reach the other team	"Always depending on the hero, but, like, almost every hero has a particular strategy you wanna follow, or separate particular strategies. And depending on the role you play your hero in, you just choose this strategy" (32DO)
	Flexible Precursory Strategy	The player has a specific strategy, but it is flexible depending on the roles of the players of the team	"I'm not really keen on playing a specific role, so I will always be the last one to pick, so that's a strategy I have, I guess" (12LO)
	In-Vivo Strategy	The player creates a strategy on the spot	"I get information out of the map, and erh, think what I have to do, or what I should do, and that's it" (25DO)

(continued)

Table 15 (continued)

Strategy making	Inductive category	Description	Example
	Absence of a strategy	The player does not have any strategy or waits for other people to make a strategy and tell it to him	"Not really [giggle], not by myself, I tend to play with others, not alone, I tend to play with others and, ehm, I realise that I rely on them, but if there's a strategy necessary, I rely on them to tell me how to play. I know how to play my characters, and what to do, but, yeah. I don't really think about the strategy, a big strategy, all over the, all over the game." (37DO)
Strategy communication	Answer to the question "Do you communicate your strategy?"		
	Inductive category	Description	Example
	Written chat	The player communicates their strategy using written chat	"I would communicate it by the chat, ehm, depending on the, on the whole situation, if eh, opponents die, if eh, my team members die." (09LO)
	Voice chat	The player communicates their strategy using a voice chat	"Yeah, it's usually why I chat or (...) voice in skype" (09LO)
	Pings	The player communicates their strategy using game pings	"I'm using smart, erh smart pings, to tell them that I'm coming, or the enemy is missing on the lane" (16LO)
	No communication	The player does not communicate their strategy	"I don't generally tell them" (17LO)

After this procedure, the statements separated according to the inductive categories were submitted to a process of semantic generalisation; thus, accounting for the implicit content of the statements. Afterwards, a reduction summary was made to consolidate and integrate the implicit information obtained in the generalisation process following the steps of the reduction process as described by Mayring (2014, pp. 68–77).

As an example of this process, the statement "But I check it regularly about where my team is, where the enemy is, how much of the map I can see, where I can gather, to farm, or to put wards and stuff." (27DO) was first coded under the 'seeking general information' code. In the second step the generalisation process was conducted, allowing for detecting the information the player seeks, for this specific example: The location of the team, the location of the enemy, visibility of the map, minion and monster clusters, and *warding* locations (process of putting wards in the map, see Appendix F). Afterwards, the reduction summary is made, consolidating the information acquired in the following sentence: "The player uses it to look for his team, his enemies, minions and farm, as well as the area of the map that is visible or invisible (map awareness)".

3.8.3. Eye-tracking

After the data had been gathered, predefined Areas of Interest (AoI) and gameplay sections were processed using PupilPlay (Pupil-labs, 2014) and exported to a CSV database. The database was organised according to the number of participants (N) and observations (games played) per participant (named as N_1).

The raw data was organised according to the participants, and the AoIs were processed with R v3.2.1 (Robert & Ross, 2016) using the 'Saccades' package for R (von der Malsburg, 2015) to detect fixations and saccades. This algorithm was used because it allows for detecting saccades according to the amplitude measured in the change of velocity between fixations, which proves useful for differentiating high amplitude saccades (indicators of information search) from small area saccades. The data was then compiled into a database together with the demographic data of the sample. Afterwards, the data was separated according to the sample groups and then cleaned of missing values by deleting rows (observations) where missing values were present for any of the variables. Next, the data was examined for extreme outliers and any observations above or below three standard deviations were deleted (Bolzer, Strijbos, & Fischer, 2015). The code used and a detailed explanation of the procedure can be found in the repository of Díaz (2015). After the data had been cleaned of outliers, the initial 54 observations for LoL players were reduced to 53, while the initial 50 observations for DotA2 Players were reduced to 47.

Finally, R v3.2.1 was used to examine whether (a) there is a relationship between saccades and fixations according to the levels of expertise of LoL and DotA2 players, and (b) if there is a difference between players' saccades and fixations according to the game played.

Due to the variability in the duration of game play (see Table 16) and its influence on the number of fixations and saccades, the quantitative indicators chosen for further analysis were those independent from the gameplay duration: Fixations per Second (FpS) on Area of Interest (AoI –the current study focuses on the areas game and minimap), Saccades per Second (SpS) on AoI, and the Mean duration of Fixations (MDF) on AoI.

Table 16

Variability of playing times according to the game played

Player group	N	N_1	Mean (sec)	SD (sec)	Min (sec)	Max (sec)
LoL Players	18	53	1958.23	522.18	922.70	3004.23
DotA2 Players	10	47	2469.12	741.74	1295.47	4404.30

3.9. Validity and reliability

Validity and reliability were assessed following the audit procedure by Akkerman et al., (2008) which details guidelines to evaluate quality in complex research processes. The audit procedure aims to account for confirmability (objectivity) and dependability (reliability) in four different stages of a research study: (a) Designing and writing the research proposal, (b) gathering data, (c) data analysis, results and conclusions, and (d) reporting the research. The procedure consists of documenting carefully and detailed the different stages of research, selecting an auditor for the project, and working in a circular manner to discuss any problems assessed by the auditor. Afterwards, the auditor (in this case the second author) generates a report stating the integrity of the research study. The report issued by the auditor can be seen in Appendix E.

3.10. Results

3.10.1. Qualitative analyses

The qualitative analyses presented in this section obey to the integration of the different levels of analyses conducted on the semi-structured interviews: The statistical distribution of statements by groups and deductive codes, the statistical distribution of statements by group and inductive codes, and semantic generalisation and integration of the data.

Table 17 presents the frequency of codes describing the deductive and the inductive categories and serves as a statistical reference about the differences between groups. As a note, the recurrences are presented in percentages based on the number of times a code appeared divided by the number of interviews conducted. This way of reporting the recurrences was chosen for allowing a relative depiction of the appearance of certain topics in the interviews which is more discerning when comparing groups than the total count of appearances.

Table 17

Frequencies and relative frequencies of codes used in the analyses of the interviews for LoL and DotA2 players

Deductive category	Code frequency in LoL players' interviews*	Relative code frequency according to LoL Players	Code frequency in DotA2 players' interviews*	Relative code frequency according to DotA2 Players
Hypotheses	65	361.11%	43	430.00%
Minimap	34	188.89%	20	200.00%
Strategy making	20	111.11%	17	170.00%
Strategy communication	13	72.22%	10	100.00%
Inductive category	**Code frequency in LoL players' interviews**	**Relative code frequency according to LoL Players**	**Code frequency in DotA2 players' interviews**	**Relative code frequency according to DotA2 Players**
Hypothesis generation	14	77.78%	11	110.00%
Hypothesis topic	19	105.56%	14	140.00%
Hypothesis occurrence	15	83.33%	8	80.00%
Hypothesis testing	6	33.33%	0	0.00%
Interpretation of results	6	33.33%	1	10.00%
Hypothesis backing	1	5.56%	2	20.00%
Hypothesis goal	1	5.56%	2	20.00%
Information sharing	4	22.22%	2	20.00%
Intervening factors	3	16.67	1	10.00%
Integration of hypotheses into strategy	29	161.11%	9	90.00%

(continued)

Table 17 (continued)

Inductive category	Code frequency in LoL players' interviews	Relative code frequency according to LoL Players	Code frequency in DotA2 players' interviews	Relative code frequency according to DotA2 Players
Map awareness	5	27.78%	9	90.00%
Game awareness	5	27.78%	1	10.00%
Seeking general information	12	66.67%	5	50.00%
Seeking specific information	16	88.89%	12	120.00%
Navigating and exploration	1	5.56%	2	20.00%
General Precursory Strategy	12	66.67%	2	20.00%
Specific Precursory Strategy	2	11.11%	5	50.00%
Flexible Precursory Strategy	6	33.33%	1	10.00%
In-Vivo Strategy	6	33.33%	7	70.00%
Absence of a strategy	0	0.00%	1	10.00%
Written chat	5	27.78%	0	0.00%
Voice chat	1	5.56%	3	30.00%
Pings	4	22.22%	1	10.00%
No communication	2	11.11%	0	0.00

*The code frequency only accounts for the inductive and deductive categories used in the analyses of the interviews and does not account for the implicit information found in the generalisation analyses.

Table 17 summarises the trends of topics found in the interviews on LoL and DotA2 players. These trends represent the emphasis players give to different actions and aspects of the games when playing, as stated by them in the semi-structured interviews. Concerning the deductive categories, it is possible to observe in Table 17 a trend in DotA2 players for more insightful reflection on their games, with a higher relative frequency of appearance of the deductive categories codes. Regarding the relative frequency of the inductive codes related to hypothesis making, it seems like DotA2 players tend to generate more hypotheses and more specific hypotheses than LoL players (see ‘Hypothesis generation’, ‘Hypothesis topic’, and ‘Hypothesis goal’ in Table 17). Nevertheless, LoL players tend to test their hypotheses and re-evaluate them more than DotA2 Players (see “Hypothesis testing’ and ‘Interpretation of results’ in Table 17). Additionally, it seems that DotA2 players back more their hypotheses than LoL players (mostly using previous experience, either direct or vicarious), while LoL players tend to integrate their hypotheses into their strategy more than DotA2 players.

Regarding the map awareness, DotA2 players seem to be more focused on it than LoL players, mostly for seeking specific information (see ‘Map awareness’ and ‘Seeking specific information’ in Table 17). The tendencies found for the inductive categories linked to Strategy making and Strategy communication will be explained in depth in sections 3.10.1.3., and 3.10.1.4.

3.10.1.1. Hypothesis generation

The LoL players reported that they mostly gather information about the position and behaviour of the enemies from the environment. Players can recognise patterns of action related to the position of players according to their role based on these environmental cues. Then, players hypothesise whether an action could take place due to the presence or absence of certain character/role in a determinate place and what that action might be. For instance, an absent *jungler* (see Appendix F) can mean that a *gank* (game jargon for ambush; see

Appendix F) is going to take place. Based on the interviews, LoL players make three types of hypotheses about (1) a probable action based on the absence or presence of players on the map (e.g., "Maybe if I have a ward in, on the raptor, can, I can say 'Yeah, probably the jungler is coming to bot-lane next'" (11LO)), (2) the probable outcome of a fight (e.g., "before or after a fight you think, okay how does it go down. What abilities do we have at our disposal, what do they have. What are they probably going to look for here, and how can we use that to our advantage" (04LO)), and (3) the probable location of a player (e.g., "I know he can maybe help, if something goes wrong in the lane, or to see where the enemies are, so I can expect an incoming gank either from, from one of the solo-laners, or the jungler" (08LO)).

Additionally, different hypotheses types are linked to player roles in the game; therefore, a player with a *marksman role* (see Appendix F) will hypothesise about where the enemy is located or will be moving next more often, probably because their role is based on long-distance attacks (e.g., "[you are] trying to guess where the enemy is going based on where you saw them last" (07LO)), whereas a player with a *support role* (see Appendix F) will hypothesise about possible safe zones on the map more often (for instance where to place *wards* or counter them), probably because they have to protect other players, ensure that other players can see certain areas, and prevent the enemy to *ward* the terrain (e.g., "I dunno, because I play support maybe, and I am always the one getting the pink wards and clearing the other ones" (05LO)).

The LoL players stated that they use map awareness to gather information to detect certain patterns in the game. Hypotheses based on map awareness are swift and more related to intuition -understood as probabilistic reasoning (Schum, 2001), that is, these hypotheses are not tested but rather reflect a 'gut feeling' that guides action. Based on the interviews, hypothesis generation of LoL players consists in detecting a pattern and quickly detect the most probable outcome to make a decision (see ‘seeking general information’, ‘seeking specific information’, ‘hypothesis generation’, and ‘hypothesis testing’ in Table 17). Sometimes players share their hypotheses with their teams, and if so, they typically do this via pings (see ‘strategy communication’ and ‘pings’ in Table 17).

DotA2 players reported gathering information about the enemy team via voice chat or ping communication (see ‘strategy communication’, ‘voice chat’, and ‘pings’ in Table 17). Players make hypotheses based on the information gathered. DotA2 players do not make hypotheses based on probes (as LoL players do – this is backed in the proportions of ‘hypothesis generation’ and ‘hypothesis testing’ for LoL and DotA2 players in Table 17), but what Popper (1959) calls “probability statements”. Based on the information they have, players estimate possible enemy behaviours. DotA2 players made remarks illustrating that they hypothesise about the probabilities of winning and losing when engaging in certain battles. Based on the interviews, DotA2 players make three types of hypotheses (see Table 17):(1) Probable attributes of enemy (i.e., location, equipment, and level), (2) probable behaviour(s) of the enemy (e.g., “'I'm always thinking that with that, ehm, advanch-advantage the enemy team has, they can, they can do certain things” (36DO)), (3) and probable outcomes of battles (e.g., “But lonely, I, I look at my team, and look at the enemy team and see what I can do to win, erh, yes, to maximise the chances to win that way around” (27DO)). Few players (see ‘Hypothesis backing’ in Table 17) stated backing their hypotheses and decisions in expertise and previous experience (e.g., “Because erh, it's very rare that five heroes are missing” (34DO)).

Additionally, LoL and DotA2 players make different types of hypotheses. LoL players tend to play based on roles, so their hypothesis systems are normally based on where players are located and what they are doing. Conversely, DotA2 players make hypotheses based on their individual and particular strategies, trying to adapt them according to the necessities.

This makes that DotA2 players centre their hypotheses in what is convenient for the short and long term strategies.

3.10.1.2. Use of the minimap

The LoL players reported that they look at the minimap mostly to locate the enemy (particularly the *jungler* for its role as *gank initiator* -see Appendix F) (e.g., “To just be aware of the situation, if there is an enemy missing, for example, you can still with a quick look on the minimap, or if you actually have some wards up- actually if you have some wards up, it is nice information to get from the minimap. If you have them up, you can check if you have some vision of the enemy jungler, for example, or if they have some heavy roamers it's important to keep track of that.”(03LO)); they also look for where their teammates are and if they are in trouble (e.g., “I'm looking at, erh, or where my hero is, erh, moving. And erh, I don't know if my teammate is in trouble, or if the enemies has shown up” (25DO)). Players pay particular attention to off-set situations, for instance, players that appear ‘absent’ on the map where there should be someone might indicate the deployment of a certain strategy by the enemy team. Some players use the minimap to gather other types of information, such as their team or *enemy minions*, this is also referred to by the players as *map awareness* (see Appendix F). According to the implicit content analysis, few of the LoL players mentioned the minimap as a pivotal point for strategy creation; on the contrary, players rely more on patterns of interaction that emerge during the gameplay (game awareness). (See ‘map awareness’ and ‘game awareness’ in Table 17).

For DotA2 players, the minimap is the cornerstone of hypotheses, and hypotheses are the cornerstone of strategy. The minimap allows for an overview of the game and provides players information (i.e., the location of the enemies, their items, and behaviour) that they can use to form their hypotheses. It also provides information about the disposition and behaviour of the team, e.g., the location of the *creeps* (see Appendix F), of the enemy (opportunity to *farm gold* and *experience*, see Appendix F), and the creeps of the team (opportunity to attack *towers* or indication of them losing territory). This information is important to make and prevent *ganks*. DotA2 Players also look at cues such as *monster camps*, *Roshan*, and the enemy *jungler*. Based on the interviews it is possible to say that map awareness is the key element in DotA2 player's strategy generation (See 'map awareness' and 'game awareness' in Table 17).

3.10.1.3. Strategy making

Most of LoL players stated to have a strategy to play the game. Four strategies could be identified. The 'General Precursory Strategy' consists of knowing some patterns from the game, such as roles within the game and game phases and trying to adapt to these patterns depending on how the team develops. The 'Specific Precursory Strategy' consists of organising the team and the players' own game character taking into consideration as many pieces of information as possible from the beginning of the game (the position and role of the team heroes, the position of the enemy, advantages, disadvantages, equipment, and game phases). The 'Flexible Precursory Strategy' is when players have a specific strategy for playing the game (e.g., a specific position in the team or a specific *Hero built* -see Appendix F) but they are open to change this strategy to adapt to the necessities of their team. The 'In-Vivo Strategy' refers to players who do not have a defined strategy, but instead develop a strategy according to the situation during the game by adapting to the needs of the team, contributing information, generate hypotheses, and mending the errors made in the *early game* (see 'Game phases' in Appendix F). The In-Vivo Strategy obeys to the necessity of the team to adapt to the constant changes in the status quo of the game. The final observation was the ‘Absence of a strategy’ and players without a predefined strategy who relied on instructions by other team members. The LoL players also pointed out that map awareness, enemy behaviour, and items are important pieces of information needed to generate their strategies. They mostly used a combination of a ‘General Precursory Strategy’ (e.g., how to situate a character in a lane according to their role), ‘Flexible Precursory Strategy’ and ‘In-Vivo Strategy’ (i.e., solving issues as they arise) (See Table 17). Nearly all the interviewed LoL players stated that they did not have specific strategies (i.e., ‘Specific Precursory Strategy’) because they just wanted to have fun and enjoy the game (e.g., “I try to win the game, but I mainly focus on having fun, while playing the game, and try that my team also has fun. I, yeah, I try to support them, and, yeah I. That's basically it”, (11LO)). Only two players stated to have a predefined strategy.

The same four types of strategy were found for DotA2 players. DotA2 players prefer using ‘Specific Precursory Strategies’ and ‘In-Vivo Strategies’ (see Table 6) by planning different roles and strategies, and reviewing different characters builds (e.g., “And second game, I picked Enigma because ehm, it's, especially when you playing team against team, will, when people gather together and has real team fights, eh, Enigma has an ultimate that disables all people in the area, and it might be helpful for the team” (30DO)) from the *early game*. This combination allows them to solve intermediate problems or develop counterstrategies as difficult situations arise. The DotA2 players also indicated that map awareness and enemy behaviour are important pieces of information needed to generate their strategies.

3.10.1.4. Strategy communication

When asked about their strategy communication, players of LoL did not emphasise on the strategy they communicate, but in the media they use for general communication. Players preferred pings for being faster, voice chat when playing with friends on the team (using external software) and written chat. The written chat consumes time; therefore, they try to use it only for specific circumstances (e.g., “If I play with a few friends, of course we communicate, but it's hard to do that, erh, through the [written] chat” (13LO)) or to give short, specific instructions (e.g., “Sometimes I'm also writing in chat that ‘TP’ is up soon, or that I'm going to ‘TP’ to them, if they place a ward” (16LO)). LoL players also emphasised the social aspect of the game, regarding sportsmanship as an important aspect (i.e., say "hello" or "have fun" at the beginning of the game).

The DotA2 players conveyed that communication is the key to victory throughout the game from role assignment to the particular procedures during gameplay. Nevertheless, they also highlighted that communication within the game could be problematic. Some DotA2 players stressed the problem to be on themselves because they do not want to communicate, or communicate very late in the game when some things have been tacitly decided in the beginning (e.g., “Ehm, at the beginning of the game, uhm, no, but thinking about it, it would be better, because sometimes the other h-uhm-players in your team don't understand what you're doing, so it would be good if everyone is, ehm, sharing the same strategies at the start of the game. But, ehm, yeah I'm mostly doing it right when it fail, so afterward's just like; ‘Oh no, I wanted t-to have it like this, and it happened like that’ so...bad [loud tsk]” (36DO)). Other players stressed that communication is often impeded by players who do not speak English, as it is clearly impossible to communicate (e.g., “Yes! I try to at least. Because it's - in DotA - it's normally, like, there're communication issues, uhm, because lots of Russians play the game –giggle-, erh, uh, either they're not very good, or they don't speak any English at all” (27DO)). As for the preferred medium of communication, none of the DotA2 players stated that they used the written chat. Instead they prefer voice chat as it enables very fast communication of important information. Pings were also important for DotA2 players, but not as important as they were for LoL players (See Table 17).

3.10.2. Comparison between LoL and DotA2.

3.10.2.1. Hypothesis generation

LoL and DotA2 players make different hypotheses, both in type and form. LoL players tend to play based on roles; therefore, their hypothesis systems are also linked to their roles. Their hypothesis systems are also simpler and based on pattern recognition. LoL players limit their hypotheses to the patterns detected and base their actions on the most probable action, least dangerous outcome, or most effective procedure. For Hypotheses generation, players rely on game awareness, which is more appropriate for detecting patterns of action and enemy strategies. Their strategies are mostly general and, although they can adapt their strategy for short-term problem solving, their general strategy is highly linked to the role performed.

In contrast, hypotheses seem to be the cornerstone for DotA2 players' strategy making. Hypotheses by DotA2 players are made following a series of probabilities that are strongly based on experience playing the game and information exchange inside game communities (for an insight on the relevance of gaming communities for knowledge building in games see Gee, 2007a; Steinkuehler, 2008, 2010; Williams, 2015; Young et al., 2012), and gathering information from the current game (map awareness).

DotA2 players' probabilistic hypotheses open the door for several sets of possible strategies and counter-strategies that can be implemented during the game as short term (e.g., survive a *gank*) or long term (e.g., win the game) problem solving

Players of both games have in common that they back their hypotheses with previous experiences, either from their gameplay or gaming communities (e.g., watching professionals playing).

3.10.2.2. Minimap

The minimap is used differently by players of both games. For LoL players, the minimap acts as the main source for game awareness and allows them to detect patterns that act as indicators of further possible actions in the game. In contrast, DotA2 players try to gather as much information as possible from different sources (i.e., minimap, sound cues, chat, information exchange with the team, and stats window) for accurately formulate their hypotheses. Whereas LoL players focus on specific elements (like the position and movement of the *jungler*) allowing them to detect emergent patterns required for testing their hypotheses, DotA2 players triangulate information from different sources for detecting enemy play strategies, using the minimap as the cornerstone for their information seeking strategies.

3.10.2.3. Strategy making

Four types of strategies were found for both DotA2 and LoL players ‘General Precursory Strategy’, ‘Specific Precursory Strategy’, ‘Flexible Precursory Strategy’, and ‘In-Vivo Strategy’. Nevertheless, players of both games used these strategies differently.

On the one hand, LoL players preferred General Precursory Strategies based on the role they played and mixed them with Flexible and In-Vivo Strategies to solve problems or deal with unexpected situations during the game. DotA2 players, on the other hand, were more inclined to use a Fixed Previous Strategy combined with In-Vivo Strategies. This strategy combination allows DotA2 players to set long term goals in problem solving while enabling them to solve problems that arise during the game.

Additionally, LoL players tend to integrate their hypotheses into their strategy more than DotA2 players, this is probably because LoL players strategies are more flexible, thus allowing for changes result of the generation and evaluation of new hypotheses compared (see ‘Interpretation of results’ and ‘Integration of hypotheses into strategy’ in Table 17). As DotA2 players are more long-term strategy oriented, this behaviour, although important, is not as prevalent as with LoL players.

3.10.2.4. Strategy communication

DotA2 and LoL players report the role of communication in the game play differently. LoL players focus on the "communication" part, emphasising the media used for communication over the strategies used. They also emphasise the fun and social elements of the game, focusing more on this than in a winning strategy. Sportsmanship, like saying "hello", "have fun", and "*gg*" (Good Game) is valued by them. They also prefer to communicate with pings and with an external voice chat (i.e., TeamSpeak) if they are playing with friends.

DotA2 players focus on the "strategy" aspect of their communication, emphasising the importance of communication for winning and overcoming difficult situations. Communication of strategies already happens at the beginning of the game to establish roles and *built* of the characters. They also emphasise problems in communication due to the lack of knowledge of the language and how this affects the gameplay. When it comes to the communication part, players prefer to use voice chat and pings.

3.10.3. Quantitative analyses

3.10.3.1. Descriptive analyses

Descriptive statistics for the eye-tracking data were computed with the package "Psych" v1.4.3 for R. Table 18 depicts the descriptive statistics for fixations and saccades for both AoIs (game and minimap) separately according to the game played and aggregated for both games combined. The quantitative analysis does not aim study differences in players' expertise, but to detect differences and similarities in the SR and SM patterns used by players of different games and levels; this is conducted under the premise that some patterns can only be observed by contrasting different player types (Bavelier et al., 2012). Furthermore, the quantitative analysis aims to support the qualitative findings of the study.

The analyses were conducted using the observations (N_1) as sample. Analyses were done using six eye-tracking variables as the dependent variables: FpS (game), FpS (minimap), SpS (game), SpS (minimap), MDF (game) in seconds, MDF (minimap) in seconds. The independent variables were the different levels of expertise of the players of both games (A to H for LoL, and I to N for DotA2).

Table 18

Descriptive statistics for fixations and saccades by AoI, and each game separately as well as both games combined

LoL Players					
Variable	N_1	Mean	SD	Min	Max
FpS (game)	47	1.35	0.41	0.44	2.30
FpS (minimap)	47	0.07	0.05	0.00	0.21
SpS (game)	47	10.05	2.91	3.38	15.27
SpS (minimap)	47	0.50	0.35	0.02	1.46
MDF (game) in seconds	47	0.25	0.04	0.20	0.37
MDF (minimap) in seconds	47	1.54	2.11	0.15	9.18
DotA2 Players					
FpS (game)	44	1.44	0.25	0.71	1.87
FpS (minimap)	44	0.11	0.11	0.00	0.29
SpS (game)	44	10.78	1.61	4.96	13.09
SpS (minimap)	44	0.81	0.77	0.02	2.23
MDF (game) in seconds	44	0.24	0.02	0.22	0.31
MDF (minimap) in seconds	44	1.23	0.31	0.12	8.05
LoL and DotA2 players					
FpS (game)	91	1.39	0.36	0.44	2.30
FpS (minimap)	91	0.09	0.08	0.00	0.29
SpS (game)	91	10.41	2.59	3.38	15.27
SpS (minimap)	91	0.65	0.55	0.02	2.23
MDF (game) in seconds	91	0.24	0.03	0.20	0.37
MDF (minimap) in seconds	91	1.39	1.96	0.12	9.18

3.10.3.2. Differential analyses

Analyses of Variance (MANOVA) was performed on the difference between players' categories of expertise, first according to each of the games and then for both games according to the combination of the six eye-tracking variables studied. Medium statistical differences were found for the within-group analyses for LoL and DotA2 players ($p<.05$). Strong statistical differences were found for the between-group analyses for LoL and DotA2 players ($p<.001$). Table 19 summarises the findings.

Table 19

MANOVA within-group analyses for LoL and DotA2 players and MANOVA between-group analyses for both LoL and DotA2 players

	df	Wilks' λ	F	sig	η^2
LoL Players (within group analysis)[1]	4	.364	1.824	.017**	.113
Error (LoL Players within group analysis)	42				
DotA2 Players (within group analysis)[1]	4	.284	2.168	.003**	.139
Error (DotA2 Players within group analysis)	39				
LoL and DotA2 Players (between group analysis)[2]	9	.304	1.910	.000***	.157
Error (LoL and DotA2 Players between group analysis)	81				

Note. * p < .10, ** p < .05, *** p < .001
[1]3 observations deleted due to missingness
[2]6 observations deleted due to missingness

In Table 19 it is possible to observe how the analyses within-group for LoL Players and DotA2 players possess statistical difference ($p<.05$), suggesting a difference between the groups evaluated according to the variables analysed. There is also a statistically significant difference in the analysis comparing the groups of LoL and DotA2 players ($p<.001$) indicating a strong difference between the groups according to the variables analysed. Moreover, the within-group analysis for LoL shows a medium effect size ($\eta^2<.14$) indicating a medium practical difference between groups; while the within-group analysis for DotA2 and the between group analysis depicted a large effect size (η^2>.14) indicating a large practical difference between the groups according to the variables analysed.

Following, Analyses of Variance (ANOVA) were made for testing differences between the groups of expertise according to the six eye-tracking variables analysed. With this, we aimed to detect the influence of particular variables in the results of the MANOVA analyses while correcting for multiple comparisons using Bonferroni test. It was found that the LoL groups are very homogenous regarding the six eye-tracking variables studied, while the DotA2 groups differ more. It was also found a significant difference between LoL and DotA2 players. Table 20 summarises the findings of the ANOVA analyses. Additionally, the compared distribution of the groups of players regarding the variables analysed for both games are graphically depicted in Figures 9 to 14.

Table 20

ANOVA within group analyses for LoL and DotA2 players and ANOVA between group analyses for both LoL and DotA2 players

Variable	Sum of Squares	df	Mean Square	F	sig	η^2
LoL Players						
FpS (game)	.909	4	.227	1.412	.247	.089
Error (FpS -game)	6.766	42	.161			
FpS (minimap)	.019	4	.005	1.998	.112	.122
Error (FpS -minimap)	.099	42	.002			
SpS (game)	67.350	4	16.839	2.195	.086*	.133
Error (SpS -game)	322.200	42	7.671			
SpS (minimap)	.707	4	.177	1.521	.213	.096
Error (SpS -minimap)	4.875	42	.116			
MDF (game) in seconds	.003	4	.001	.631	.643	.042
Error (MDF –game)	.053	42	.001			
MDF (minimap) in seconds	21.693	4	5.423	1.242	.308	.079
Error (MDF -minimap)	183.373	42	4.366			
DotA2 Players						
FpS (game)	1.134	4	.283	4.258	.006**	.242
Error (FpS -game)	2.5956	39	.066			
FpS (minimap)	.079	4	.019	2.505	.058*	.158
Error (FpS -minimap)	.306	39	.008			
SpS (game)	58.598	4	14.149	3.959	.009**	.229
Error (SpS -game)	144.327	39	3.701			
SpS (minimap)	4.139	4	1.035	2.570	.052*	.161
Error (SpS -minimap)	15.700	39	.403			
MDF (game) in seconds	.003	4	.001	1.915	.391	.126
Error (MDF –game)	.014	39	.000			
MDF (minimap) in seconds	13.498	4	3.375	1.056	.391	.073
Error (MDF -minimap)	124.603	39	3.195			

(continued)

Table 20 (continued)

LoL and DotA2 players

Variable	Sum of Squares	df	Mean Square	F	sig	η^2
FpS (game)	2.206	9	.245	2.121	.037**	.171
Error (FpS -game)	9.361	81	.116			
FpS (minimap)	.136	9	.015	3.034	.004**	.229
Error (FpS -minimap)	.405	81	.005			
SpS (game)	138.070	9	15.341	2.664	.009**	.206
Error (SpS -game)	466.530	81	5.759			
SpS (minimap)	7.017	9	.779	3.069	.003**	.230
Error (SpS -minimap)	20.576	81	.254			
MDF (game) in seconds	.009	9	.001	1.147	.340	.101
Error (MDF –game)	.068	81	.001			
MDF (minimap) in seconds	37.305	9	4.145	1.090	.379	.096
Error (MDF -minimap)	307.977	81	3.802			

Note. * p < .10, ** p < .05, *** p < .001

In Table 20 it is possible to observe that the only statistically significant difference between the players analysed for the LoL player was found in the variable SpS (game) (p<.10), indicating a small statistical difference for players' groups regarding this variable but a general homogeneity of players' groups amongst the other variables analysed. Additionally, it was found a medium effect size for the SpS (game) variable (η^2<.14), indicating a medium practical difference between the players' groups according to this variable. For DotA2 players small and medium statistical significant differences were found for all the variables (p<.10 and p<.05) except for the variables MDF(game) and MDF(minimap) which difference was not significant, indicating that players' group variability for the all the variables except for MDF (game) and MDF(minimap) which remain homogeneous amongst players' groups. Also, a large effect size was found for all the statistically significant variables (η^2>.14), indicating a large practical difference between the players' groups in the statistically significant variables. For the between-group analyses medium statistical significant differences were found for all the variables ($p<.05$) except for the variables MDF(game) and MDF(minimap) which difference was not significant, indicating that players' group variability for the all the variables except for MDF (game) and MDF(minimap) which remain homogeneous amongst players' groups. Additionally, a large effect size was found for all the statistically significant variables ($\eta^2>.14$), indicating a large practical difference between the players' groups in the statistically significant variables.

Figure 9. Comparative depiction of the distribution of Fixations per Second (game) according to the group of expertise for LoL and DotA2 players

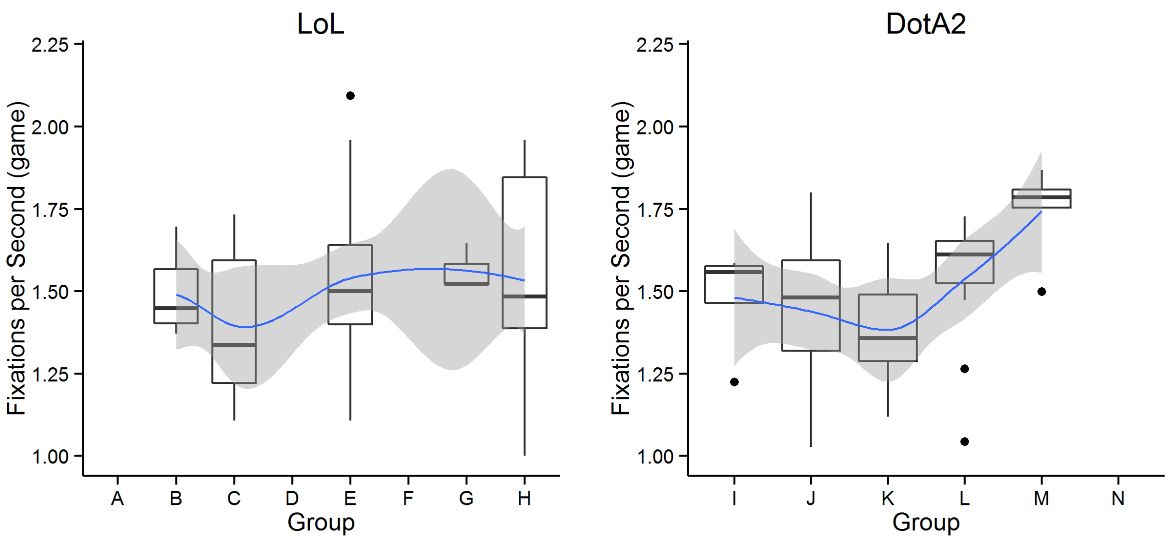
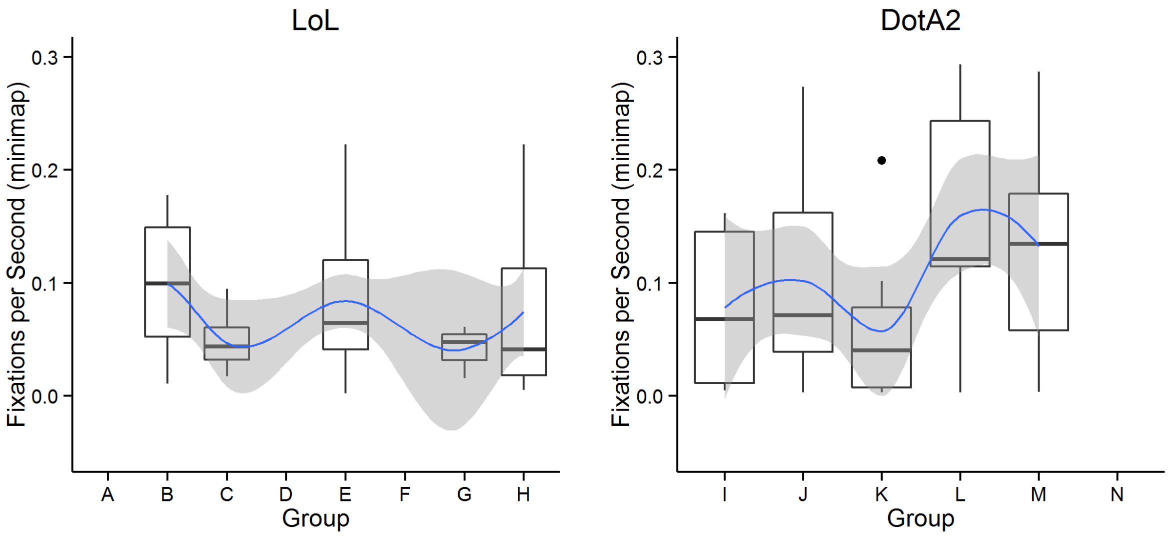


Figure 10. Comparative depiction of the distribution of Fixations per Second (minimap) game according to the group of expertise for LoL and DotA2 players



In Figure 9 and Figure 10, it is possible to graphically observe the behavioural difference for the variable FpS for both LoL and DotA2 players in the game and the minimap according to their group of expertise. It is possible to observe how in both graphs the LoL group has a more homogenous behaviour compared to the DotA2 group. Additionally, it is possible to observe an increment of the behaviour for higher groups of DotA2 players in both graphs. As indicated in section 3.7., FpS is a behaviour related to problem identification, hypothesis generation, and evidence evaluation; therefore, it is possible to infer from the graphs that while LoL players have a homogeneous use of these actions according to the groups of expertise, which agrees with the ANOVA analyses depicted in Table 20. Conversely, DotA2 players depict more variation of the behaviours related to FpS, with a decrement of these in the intermediate groups and a rise in the higher levels.

Figure 11. Comparative depiction of the distribution of Saccades per Second (game) according to the group of expertise for LoL and DotA2 players

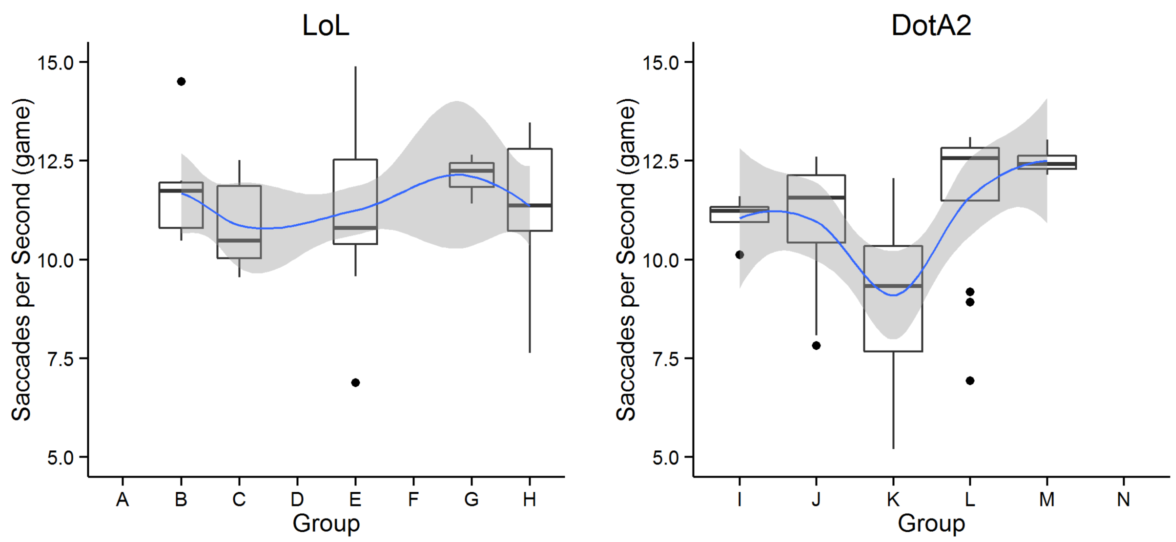
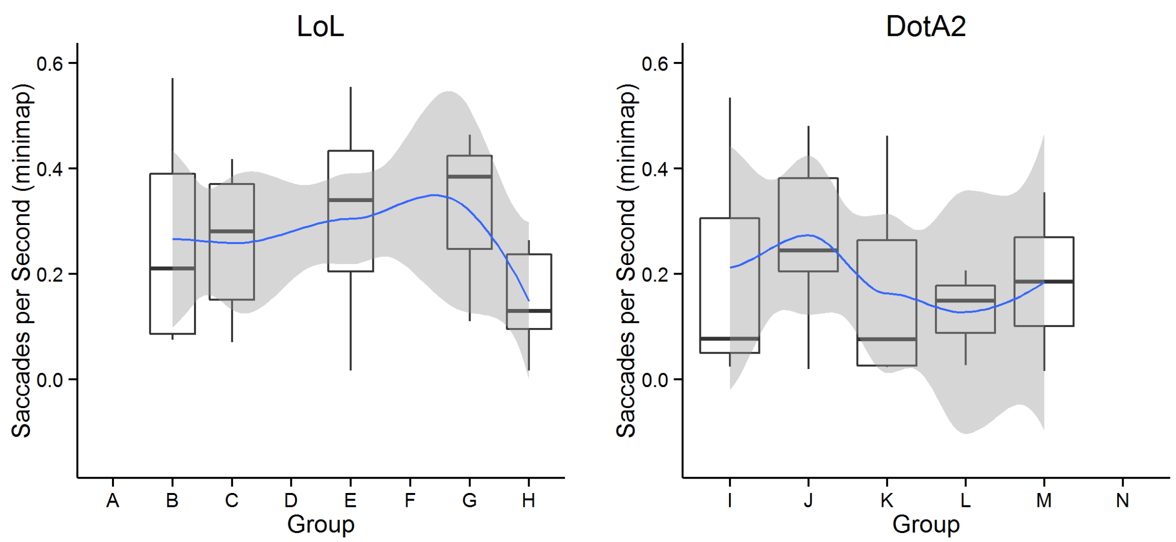


Figure 12. Comparative depiction of the distribution of Saccades per Second (minimap) according to the group of expertise for LoL and DotA2 players



In Figure 11 and Figure 12, it is possible to graphically observe the behavioural difference for the variable SpS for both LoL and DotA2 players in the game and the minimap according to their group of expertise. It is possible to observe how in both graphs the LoL group has a more homogenous tendency with a drop in the end, while the DotA2 group present a curve with a depression in the middle levels of expertise similar to those observed in Figure 9 and Figure 10. As indicated in section 3.7., SpS is a behaviour related to problem identification, information search, map awareness, game awareness, and evidence generation; therefore, it is possible to infer from the graphs that while LoL players have a homogeneous use of these actions according to the groups of expertise. Nevertheless, this behaviour within the game tends to increment with the expertise according to the ANOVA analyses depicted in Table 20. Apparently, this behaviour is opposite when related to the minimap. Conversely, DotA2 players depict more variation of the behaviours related to SpS, with a decrement of these in the intermediate groups and a rise in the higher levels, just as it does with the FpS.

Figure 13. Comparative depiction of the distribution of Mean Duration of Fixations (game) according to the group of expertise for LoL and DotA2 players

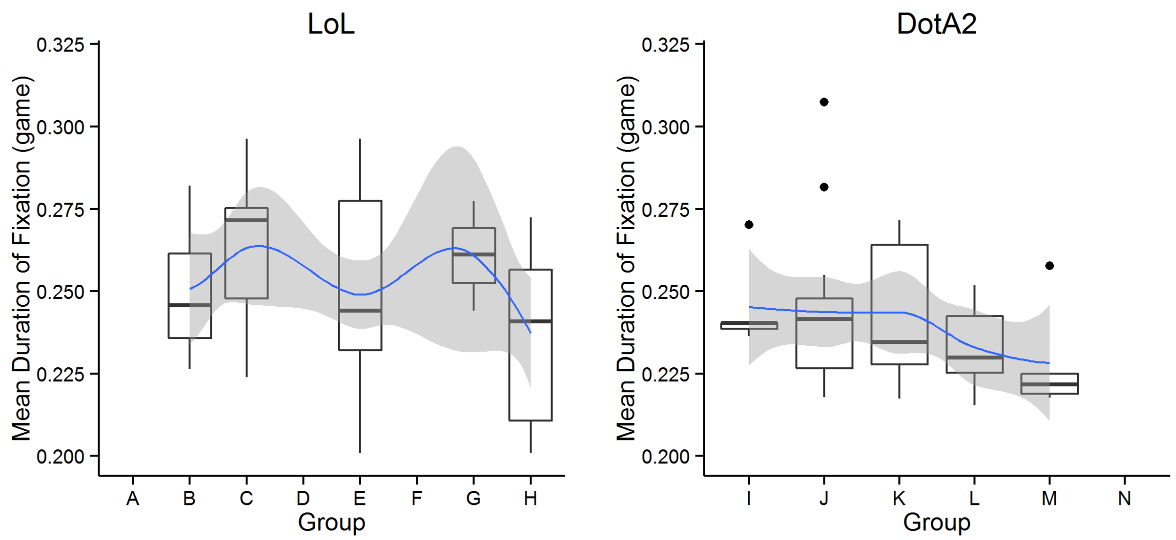
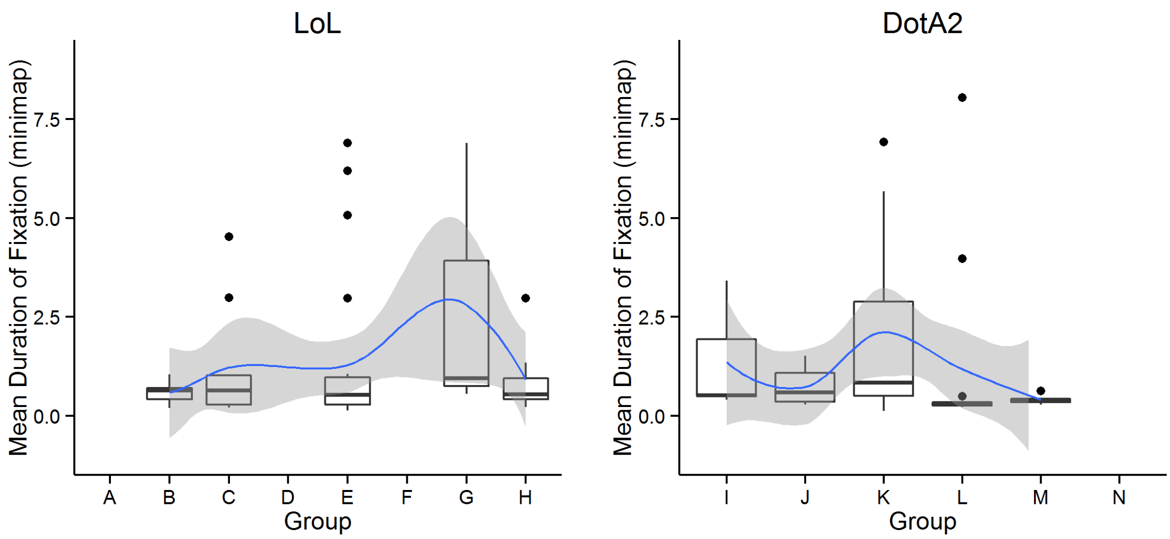


Figure 14. Comparative depiction of the distribution of Mean Duration of Fixations (minimap) according to the group of expertise for LoL and DotA2 players



In Figure 13 and Figure 14, it is possible to graphically observe the behavioural difference for the variable MDS for both LoL and DotA2 players in the game and the minimap according to their group of expertise. It is possible to observe how in both graphs the groups for both LoL and DotA2 are homogeneous. Nonetheless, DotA2 players depict a lower average amongst groups than LoL players. Although for the LoL group the variation of this behaviour appears to fluctuate, for the DotA2 group it tends to describe a curve opposite to the ones depicted in the previous graphs. As indicated in section 3.7., MDF behaviour is related to the time required for identifying new information; therefore, it is possible to infer from the graphs that although homogeneous (no statistical difference found in the ANOVA analyses) players of DotA2 tend to require less time to identify information as their expertise increases.

Although both graphs illustrate the findings of the ANOVA analyses presented in Table 20, it is possible to observe that there is a trend in DotA2 players regarding learning curve and expertise style. While LoL players depict less variation or increment of behaviours in their expertise levels, DotA2 players have a strong trend to display the behaviour in the lower levels, suppress it in the intermediate levels, and increasingly display it again in the high levels. This way of expertise development is similar to the U-shape development models described by Morse, Belpaeme, Cangelosi, and Floccia (2011).

3.11. Analysis

This section aims to analyse and integrate the qualitative and quantitative findings of the study regarding players SR and SM processes used while playing LoL and DotA2. The analysis is guided by the Research Questions: (RQ4) How MOBA players use SR and SM when playing, (RQ5) whether differences in the level of expertise of players delineate patterns SR and SM on their play, (RQ6) whether there is a difference in both SR and SM according to the game played, and (RQ7) the relationship between SR and SM when playing MOBAs.

3.11.1. Differences in levels of expertise in LoL and DotA2

Based on the qualitative analyses, players of both games describe the same patterns of hypothesis generation, hypothesis testing, strategy making, and information gathering – although the regularity of the type of pattern differs depending on whether they played LoL or DotA2.

MANOVA analyses revealed a small significant multivariate main effect for groups of expertise in the LoL group on the eye-tracking data. Follow-up ANOVA analyses revealed only a small significant difference in Saccades per Second (game) within LoL groups (Table 20) with a trend of more expert players having more saccades per second (Figure 11), thus indicating more information search.

MANOVA analyses showed a moderate significant multivariate main effect for expertise in DotA2 group on the eye-tracking data. Follow-up ANOVA analyses revealed small significant differences in FpS (minimap) and SpS (minimap), and a moderately significant difference in FpS (game) and SpS (game) within DotA2 groups (Table 20). The differences depict a trend of increasing FpS and SpS while the level of expertise increase (Figures 9 to 12). As specified in section 3.7., these differences indicate more information search and information processing in higher levels of expertise (Poole & Ball, 2005). Additionally, DotA2 players depicted fewer Fixations per Second (FpS) and Saccades per Second (SpS) on intermediate levels creating a valley effect (Figures 9 to 12).

Quantitative findings complement the qualitative findings in the extent that LoL players make simple strategies based on patterns recognition, thus making the information search (SpS) and the information processing (FpS) homogeneous (Table 20). The interviews revealed that the DotA2 players are confronted with more and fuzzier information they need to gather and triangulate to make hypotheses; this was also visible in the results of the eye-tracking analyses as a difference on information search and information processing (Table 20).

The same findings as found for the game AoI were found for the minimap AoI, where LoL players showed homogenous information search and processing behaviour for the minimap AoI with no statistical difference in FpS and SpS according to expertise levels. DotA2 players showed statistical differences in the FpS and SpS according to their expertise level, with higher levels depicting more FpS and SpS (Figure 10 and Figure 12).

3.11.2. Differences between LoL and DotA2 players

The interviews revealed large differences on how LoL and DotA2 players played MOBAs for the qualitative and quantitative analyses. LoL players' strategic cornerstone was not hypothesis generation; instead, they relied on pattern detection given the information presented in the game (game awareness). DotA2 players made complex probabilistic hypotheses about possible situations and their outcomes based on map awareness. These differences were also found in the MANOVA and ANOVA analyses for FpS (information processing) and SpS (information search). In that way, the LoL players have a general way of playing, thus the lack of differences in information search and information processing between groups. DotA2 players showed differences according to different levels of expertise on both information processing and information search (Table 20).

Although LoL players regard the minimap as a good way to acquire certain information related to the patterns they require for creating a strategy, no statistical differences were found in the ANOVA analyses (Table 20). DotA2 players, regard the minimap as an important source of information, focusing more on the minimap because this information is more crucial for their probabilistic hypotheses, this is supported by the statistical differences found in the ANOVA analyses.

The eye-tracking data revealed statistically significant differences in the MANOVA analysis between the levels of expertise of LoL and DotA2 players (Table 19). Additional ANOVA analyses revealed statistically significant differences in the information processing and information search between LoL and DotA2 players (Table 20), with a trend in DotA2 players towards more information processing and information search (Figures 9 to 12). These findings support the qualitative findings in the extent that DotA2 players need to account for more variables than LoL players when making a decision. Due to the entropy of information, DotA2 players require intense and extensive information search as well as information organisation to reduce the chaos of the information, allowing them to create and communicate strategies.

There were also found statistical differences in how both groups of players look at the minimap considering the duration of saccades and fixations, with LoL players spending less time fixating in the minimap, but more time looking for information in the minimap compared to DotA2 players (Figure 10 and Figure 12). This complements the qualitative findings as LoL players search for information allowing them to detect patterns, while DotA2 players spend more time integrating the information presented in the minimap.

The average time spent identifying information (MDF) does not show a statistically significant difference for either of the groups nor the AoIs.

3.12. Discussion

In this section, we discuss the findings in the light of the Research Questions: (RQ4) Explore how MOBA players use SR and SM when playing, (RQ5) explore whether differences in the level of expertise of players delineate patterns SR and SM on their play, (RQ6) explore whether there is a difference in both SR and SM according to the game played, and (RQ7) the relationship between SR and SM when playing MOBAs.

Accounting for RQ4, we found that the MOBA players participating in the study make use of SR and SM and that the way they make use of them is not much related to their level of expertise but on the cognitive demands of the video game played. Based on the interviews we found that players use SR processes following all the epistemic activities described by Fischer et al. (2014) while integrating these processes with SM as proposed by Miller (1989). This behaviour occurs because the game works as a problem solving environment with a big long-term goal (i.e., win the game) and smaller emergent problems with short-term goals (e.g., win a fight). This game structure fosters the parallel deployment of SR and SM processes (Blumberg et al., 2013; Jørgensen, 2003). It is important to highlight that these SR and SM processes happen very fast and non-sequentially; instead, parts of the processes (e.g., hypothesis generation and evidence generation) can be repeated until the conditions are favourable to proceed to another step of the process (e.g., communication of results), or take a step back to a previous part of the process (e.g., problem identification).

Regarding the relationship between different levels of expertise in SR and SM processes (RQ5), we found no indicators in the interviews of differences dependent on expertise. Quantitative findings indicate that players of LoL look at the game and the minimap on the same way disregarding the level of expertise, indicating the same amount of information search and information processing in the LoL players participating in the study. DotA2 players, on the contrary, depict a difference on the way they look at the game and the minimap according to their level of expertise, with more information search and processing in higher levels of expertise. These findings contrast with research on expertise with the help of eye-tracking (e.g., Dogusoy-Taylana & Cagiltay, 2014; Law, Atkins, Kirkpatrick, & Lomax, 2004) in the sense that experts normally require fewer fixations and saccades than novices when performing a task. Nevertheless, these studies were performed with problem situations in which the information is static and with a limited number of patterns. It seems that when dealing dynamic information, expert gamers behave more like novices analysing static information in the sense that they search more for information and are required to spend more effort to process the new information. It is also important to note that eye-tracking studies are typically based on static figures, whereas the present study introduced information processing and search in a dynamic situation.

The MDF, indicator of information identification, was not significant in any of the eye tracking analyses either the within-group or the between-group comparisons. This finding is expected as all players are *ranked players* (see Appendix F), and the average time they should require to process information is subject to neurocognitive structures, thus creating a ceiling effect. The average time taken by the players identifying information is 0.24 seconds which corresponds to the average time required to identify information based on eye-tracking data; this is about 0.23 seconds as suggested by studies conducted by Just and Carpenter (1976, p. 448). Furthermore, no differences in MDF were found between the LoL and DotA2 players.

Regarding the differences between players of both games (RQ6), we observed both qualitative and quantitative differences, with LoL players using SR based on game awareness and pattern recognition and DotA2 players SR based on map awareness and construction of probabilistic hypothesis. We also found that players of the two games prefer different types of strategy for solving long-term and short-term problems. These differences seem to lie in two factors; first, the differences between the games affect the way players relate to the game regarding information search and information processing. For example, DotA2 players make more complex hypotheses (based on the amount of information and dynamics they encounter through the match) and find more entropy in the game the more they advance in the ranking system – pushing the limits of the players' information search and processing capacities. This attitude towards entropy also explains why there is a statistical difference in the information search and information processing according to DotA2 players' expertise. The second factor is player's intentions; in this regard, the interviews shed light over the competitiveness of DotA2 players compared to LoL players who seem to enjoy the fun and sportsmanship of the game, thus lowering the demands of the gameplay, which could explain why the information search and information processing is homogenous. Certainly, both groups of players want to win the match and rank higher in the game, but while this is the main goal for DotA2 players, for LoL players it is not as central as it is having fun.

Players of both games at different levels of expertise mentioned in the interviews all the components of executive function described by Gilbert and Burgess (2008) and Rabbitt

(1997) (i.e., (a) doing analysis, (b) planning, (c) using prospective memory, (d) monitoring actions, (e) control an inhibition of impulses) as the game can be regarded as a non-routinely or ill-structured problem (Jonassen, 1997). For example, when a choice is to be made in the game the player (a) looks for different possibilities of actions and outcomes before making a decision, (b) thinks of a strategy, anticipating the situation even before the game has started, (c) plans how to build their character and which items are needed to do this, (d) coordinates with the team in terms of skills and actions that the player and other team members need to undertake, and (e) waits for the appropriate moment to deploy the strategy. Other components of executive function such as attention, separation of irrelevant stimuli, and multitasking were not explicitly mentioned in the interviews and could not be accounted for with the eye-tracking data. Nevertheless, research on executive function and video games indicate executive function is basic for learning and play video games (Boot et al., 2008; Gee, 2005a, 2007b; Irons, Remington, & McLean, 2011).

Regarding SR and SM (RQ7), the types of hypotheses reported by players of DotA2, compared to the LoL players, t are very complex and interlinked with their overall strategy. Their hypothesis system is the cornerstone of their strategy. The DotA2 players' stated that they build their hypotheses on information gathered from the game and the minimap. Specific items and abilities supporting information gathering (for instance items such as *wards* enable players to see through the *Fog of War* and examine hidden zones on the map) are important for the map awareness, and consequently support hypothesis generation (e.g., "But I check it regularly about where my team is, where the enemy is, how much of the map I can see, where I can gather, to farm, or to put wards and stuff" (27DO)). According to the interviews, the minimap provides players information on the number of enemies, their position, their items, and *runes* (a type of temporal power boost that can be found in DotA2 –see Appendix F), this information proves essential for map and game awareness, allowing the construction of hypothesis systems and strategies. DotA2 players stated Information regarding enemy characters and enemy items are key aspects for map and game awareness because they determine the reach of their attack and defence as well as probable strategies of the team by providing the data allowing building a hypothesis system (e.g., "The la-middle or late game, your lanes are pushing out hard, and the enemies are not responding to them, if your creeps are pushing their base, and the enemy is not pushing them out, it means that the enemy is most probably smoked and trying to find you somewhere in the lane. So, and that usually is true" (24DO)).

3.12.1. Chess is not Shōgi: Implications of the study for Game Studies and Psychology-oriented game research

This section tackles the importance of skill transfer when researching on video games. Skill transfer is the degree to which a person can use knowledge and skills of one situation to another situation. Fields with near transfer possess more similarities between them; thus, it is easier to relate the knowledge from one field to the other. Fields with far transfer are not directly linked and require more effort from the person to relate the skills from one field to another and to switch between them (Baldwin & Ford, 1988; Barnett & Ceci, 2002, 2005). Skill transfer has been a key question when studying video games, particularly from the psychological perspective for two main reasons: First, because most of the times case-control studies in video games use games as study variables assuming that they are "similar" or "dissimilar" (Fernie & DeVries, 1990), this assumption, as discussed in this research can be wrong as even very similar games can be played differently -For a deeper view on this issue see Bartle (2015). Second, because when evaluating skills development psychologists use standardised tests to account for cognitive changes in the subjects, not knowing if the skills taught by the game are transferable (Alba-Marrugo, 2013; Glass, Maddox, & Love, 2013;

Irons et al., 2011) -An exception can be drawn for educational games which ontological construct is different from that of commercial games. The present study depicted considerable differences in patterns of reasoning, strategy making, and information processing in two very similar games. Both the qualitative (hypothesis generation, hypothesis type, strategy making, and representation of the game) and quantitative (amount of search for information and amount of information processing) analyses provide support for the fact that LoL and DotA2 are not played in the same way despite being very similar. Hence, even near transfer cannot be taken for granted despite the similarities between games. This finding indicates that when studying video games in general or the relation between gaming and cognitive abilities, it is central to ensure that the compared evaluations, associated tasks, or even video games resemble each another (increasing the likelihood of near transfer) before conducting the study. Even if two games or tasks are theoretically related (Newell, 1966; Newell & Simon, 1972; Patino, Romero, & Proulx, 2016), small differences in game mechanics and player gameplay types can decrease the likelihood or even eliminate the possibility of near transfer.

3.12.2. We look there, but not for the same reasons: Methodological reflection on research in commercial games

Research in video games, particularly from the psychological perspective focuses strongly on outcomes from the game or group comparisons with alternative tasks or tests (Facer et al., 2004; Karle, Watter, & Shedden, 2010; Oei & Patterson, 2013). Nevertheless, the findings of the present research strengthen the adequacy of mixed methods when studying video games. The methodological triangulation allowed interpreting patterns and differences found in the eye-tracking data and the interviews, concerning the way the players look at the minimap, the information they look at, their intentions, hypotheses, strategies linked to certain behaviours, and information processing. Similarly, the eye-tracking showed that players of both games look at their teammates and the enemies, but the interview revealed that they do so for different reasons associated with their SM and SR patterns. The combination of eye-tracking data and interviews enabled us to identify not only where players look, but also their intentions and reasons (regarding SM and SR) for looking at a specific area or elements of the game and minimap, as well as certain behaviours and patterns. As players obey to different patterns of action (e.g., those described by eye-tracking data) but their intentions are variable (e.g., depending on the game) one-sided generalisations cannot be drawn; therefore, mixed methods arise as the best way to holistically understand the gaming process. Mixed methods shown to be particularly good for understanding cognitive processes used during gameplay.

3.12.3. A reflection on MOBAs and the educational value of video games

Researchers have inquired about the value of video games in education and cognitive development; although several of these works are found in the field of serious games, some authors have enquired about the value of commercial video games in education and cognitive development (e.g., Gee 2005a, 2005b 2005c, 2006, 2007b; Prensky 2006). Additionally, some authors have enquired how commercial video games foster literacy, social interaction, and communities of expertise dedicated to analysing the mechanics of certain games (Steinkuehler 2008, 2010). eSports, and particularly MOBAs have not been the focus of these studies as they are a recent phenomenon. The present research sheds some light on the cognitive and educational value of MOBAs and helps to understand into which extent playing these games in informal settings improves the cognitive development of the players.

First, based on previous research (Facer et al 2004; Gee 2005b, 2006, 2007a; Griffiths 2002; Prensky 2006) and from a general perspective, it is possible to say that the game itself acts as a domain of knowledge, therefore teaching the player-specific content about the game such as its mechanics and lore, without this the gamer cannot access the game world (Gee

2005a). This specific game literacy, also for the case of MOBAs, requires (and fosters by means of practice) general cognitive skills such as attention (Green & Bavelier, 2012; Irons et al., 2011), memory (Boot, et al., 2008) and fine motor skills (Borecki, Tolstych & Pokorski, 2013; Gentile, 2011). The player is required to focus on the strategy while accounting for the specificities of its character, the team's specific formation (knowing other players and heroes *stats*), and the enemy team (knowing the strategies of the opponent team as well as the specificities of their *stats* for the enemy team). In higher levels, the player is required to know (memory) which items to use, know the items their team is using, and the items the enemies possess (how the items interact between them, with the hero they are playing, and the synergy with the team). Fine motor skills are required for playing the game correctly; players usually start by getting acquainted with the controllers of the game and the association between keys and actions within the game. The first phase of the game is to interiorise the relation key-action (and sometimes key interaction-action –which is called a combo), so they can be performed without switching attention away from the screen. Higher performance players can use shortcuts and hotkeys which require memorising not only certain command, but what they do, how to do it, or how they are performed in the game (e.g., using quick cast requires to know the range in which the attack is effective). Players also require (and foster through practice) visuospatial reasoning, which is the cognitive skill allowing the player to locate themselves in the environment, calculate distances, and read 2D maps (Fischer et al., 1994; Gagnon, 1985; Greenfield, Brannon & Lohr, 1994). Lastly, it enhances creativity as with players' expertise comes in different ways to 'build a character' and to play. Creativity comes as 'playing outside the box' by innovating in the use of items, powers, spaces, tactics, and counter-tactics.

From the findings of the present research, Multiplayer Online Battle Arenas (MOBAs) require and foster (by means of practice) player's generation of hypothesis both tactical and counter-tactical (hypotheses about what enemies will do), and spatial (hypotheses about where enemies will be), at the same time it fosters hypothesis testing based on schemata (Piaget, 1962; as cited in Singer & Revenson, 1996) and based on probabilities. Probability based hypothesis making and hypothesis testing foster, at the same time, non-formal probabilistic reasoning (Green, Pouget & Bavelier, 2010). The game also fosters information seeking in two fronts: in the game, looking for cues about the enemy team behaviour and strategy; and seeking information regarding character *building* (see Appendix F), strategies, counter strategies, and lore in communities of knowledge specialised in the game (Steinkuehler 2008, 2010). MOBAs also foster long-term and short-term strategy making, the long-term strategy making requires knowledge about the game as a whole, fostering planning and player communication; short-term strategies (created to serve to advance or surpass an impasse within the game) foster cognitive flexibility (Frasca, 2001; Jørgensen, 2003;) by requiring players to change their general strategy to adapt to different situations and team formations that can arise. Moreover, MOBA games showed to strongly foster executive function (described in Chapter 3.1.) in all its different dimensions. Planning is required from the beginning of the game both from an individual and a group perspective to create general strategies. Task and stimuli attention switch is necessary for changing strategies according to different situations and acquiring map and game awareness. The game also pushes the player to prioritise relevant stimuli (i.e., to look for specific cues in the environment) and situations (e.g., fight or run), while forcing the player to inhibit some behaviours (e.g., using a specific skill) until the team requires it or until it is not dangerous to use.

Summarising, MOBAs make use, develop, and foster through exercise the use of SR (hypothesis generation, evidence generation and evidence evaluation -for hypothesis testing, drawing conclusions and communication of results), probabilistic reasoning, information seeking, creation of communities of knowledge, strategy making, and executive function.

Thus learning and fostering of skills are carried in a stealthy way (also known as implicit learning -Ciavarro, Dobson & Goodman, 2008; Fudenberg & Levine, 1998; Seger, 1994) and in an informal setting.

3.13. Conclusions

Summarising our findings and analyses, MOBA players use SR and SM processes when playing. Although the games belong to the same genre and are very similar, it was proven that players of both games do not relate to the game in the same way. Players do not only look different at the games, but they look for different information and different reasons. The strategies preferred by the players are also different depending on the game.

Intentions and cognitive demand in the game also play a role in the quality and quantity of information search and processing according to levels of expertise, with one of the games (LoL) reporting fewer differences between levels of expertise and the other (DotA2) reporting differences between the levels of expertise, both within and between groups.

3.14. Directions for future research

3.14.1. Limitations

The first limitation was the use of only two MOBA games. We focused on only two MOBA games as these two were highly similar and popular when the study was conducted. Since the data collection for the current study other free-to-play MOBAs have been released. It would be important to evaluate similarities and differences amongst similar and dissimilar MOBAs considering the differences in cognitive demand found between two homogeneous games in the current study.

In the current study, we used eye-tracking technology allowing us to gather rough-grained data as the amount of data generated during a certain period was considerable. Evaluating the same type of actions using fine-grained data, and more Areas of Interest (AoI) could improve our understanding of the way players search for information while playing and which type of information they seek.

Although each player was involved in three to five matches and the amount of information for both qualitative and quantitative data was considerable, the sample of players was rather small for both games. As a consequence, there were not players from all possible expertise levels for each game. Further studies should consider having a better-distributed sample according to the ranks for better assessing expertise.

Another challenge inherent to the MOBA games was how to measure expertise within and between games. Each game has a different ranking system, making difficult to compare players between games. As researchers cannot access the particular game ranking scales for understanding the measurement nature, game developers and eSports associations might create a common scale to eSports allowing for comparing players independent of the game they are playing.

Finally, it will be important to conduct studies with professional teams of players to examine differences in SM and SR with ranked players and evaluate a wider range of expertise amongst the gaming population.

3.14.2. Implications of the findings

The findings of the present study have implications for SR in the extent that it does not only relates SR with SM but also explores how SR processes occur in players of MOBA games.

It has implications for the research in psychology, particularly in the psychological studies in video games, in two ways; first, because it points towards the fact that even similar games might not present near transfer, which inquires about the nature of comparative research of video games in psychology. The second way is by exploring a different way to do research in video games, which take into account the complex nature of video gaming. In the same extent, this research has implications for games studies by bridging the gap between game studies and psychological game studies by proposing a holistic and exploratory approach in research about video games.

Finally, the results of the present research have implications for eSports in different areas. First, the findings concerning the learning and performance curve in LoL and DotA2 players highlight an important issue regarding the expertise classification within games. For example, while DotA2 had a clear difference between the expertise levels regarding certain behaviours indicators of cognitive processes such as FpS and SpS, LoL players had an even distribution according to expertise levels, which questions to which extent the LoL ranking system is skill differential. Second, the findings in the present research can help to establish a standard ranking amongst eSports, allowing for standardised scales of performance between games based on cognitive skill indicators, knowledge of the game, use of tools, and cooperation. Lastly, the present research can help exploring and creating models of expertise allowing for predictors of performance in eAthletes; in other words, allowing to detect certain cognitive characteristics and performance along different levels of expertise proper of highly competitive eAthletes, facilitating the discovery of high-performance players and their training.

The results of the current study pave a new way for studying video games from a cognitive perspective by inquiring on new methods of gathering data in a more context-sensitive manner and triangulating qualitative and quantitative data in a form that can explain player's cognition in a holistic way. The results of the current study also open a new branch of understanding video games as tools for fostering cognitive development in informal settings while accounting for behaviours, intentions, ideas, previous experiences, and interaction with the game and the teammates. Besides, these studies bring insight on previous and further game studies by accounting for skill transfer between games and modes of playing games of the same genre.

Chapter 4

General discussion and conclusions

4.1. Summary of findings in the first phase of the study

In the first phase of the study, we inquired about the concepts of SR and ST, how these concepts are connected and how they diverge. Using a conceptual review on 166 texts, we found that SR and ST are not the same concept; moreover, we found an ontological difference between the constructs of Reasoning and Thinking. Additionally, we proposed a way to characterise and operationalise SR based on the work of Fischer et al. (2014), together with a way to characterise ST based on the works of Harre (2004) and Popper (1966).

Parallel to the findings derived from the main research questions, we observed that some authors attribute an implicit teleology to SR when studying the concept. Additionally, we found that some authors do not explicitly operationalise their variables when conducting studies; this was particularly unexpected in the case of Empirical Research. Also, we discovered that authors characterise the concepts of SR and the Scientific Method similarly. The discussion about SR and the Scientific Method led to the conclusion that both constructs are comparable, but their domains are different as SR is seen as the set of cognitive abilities required for making use of the Scientific Method. Conversely, the Scientific method is understood as a series of steps followed for inquiring and understanding the nature of phenomena.

4.2. Summary of findings in the second phase of the study

In the second phase of the study, we inquired about how MOBA players make use of SR and SM when playing. Using a mixed method design (triangulating qualitative findings from interviews and quantitative findings from eye-tracking technologies), we found that MOBA players make use of SR when playing, and SR is the basis for SM in the game.

As parallel findings, we observed that even small differences in game mechanics could lead to utterly different playing styles required for mastering the game. Differences in game mechanics influence how players use SR and SM, as well as players' progression in the game expertise. Additionally, we confirmed that mixed methods approach is the most suitable option to research commercial games while accounting for different aspects in a holistic (comprehensive) way.

4.3. Theoretical implications of the studies

The present study comprises theoretical implications for five fields: Psychology, Philosophy of Sciences, Game Studies, Learning Sciences, and Media Studies. The first set of theoretical implications regard the fields of Psychology, Philosophy of Sciences, and Game Studies. It constitutes the differentiation of the concepts of Reasoning and Thinking as well as the ones of SR and ST. The concepts of SR and ST have been mixed in the literature either by exchanging their definitions or by using them as synonyms. By proposing a base for defining and characterising SR and ST as different constructs, we advanced in a formalisation and standardisation of the concepts that can be used to inquire and research further about them in different fields. At the same time, defining SR and ST opens the door to a regularisation of the concepts between diverse disciplines, refining interdisciplinary collaboration in their study.

The second set of theoretical implications regard the fields of Psychology, Game Studies, Learning Sciences, and Media Studies. First of all, we advanced in the theoretical research of cognitive skills used by video gamers when they play; at the same time, we deepened the theoretical and methodological system for holistically (comprehensively) understand the behaviours, representations, and skills used by video gamers of a particular game genre. A second implication is the finding that games cannot be compared as if they were homologue; even games from the same genre, console, and playing perspective require different cognitive demand and generate differences between playing styles depending on their mechanics. The observation that a change in a mechanic of a game can trigger different behaviour and cognitive demand raised the question to which extent skill transfer is possible between games and other games, between games and formal setting (i.e., schools), between games and psychometric batteries, and between games and everyday life problem solving. Finally, we proved mix-methods methodology and the triangulation of qualitative player experience with quantitative measuring to be suited for holistically understanding players' cognitive skills and behaviour within game settings.

Finally, by using the construct of SR from the first phase of the study in the second phase for both the qualitative and the quantitative analyses, we demonstrated the fitness of the construct proposed by Fischer et al. (2014) for research SR in Game Studies and Psychology domains, even from different methodological approaches.

4.4. Practical implications of the studies

The present study comprises practical implications for three fields: Psychology, Game Studies, and Learning Sciences. The first set of practical implications regard the fields of Psychology and Learning Sciences. Making the distinction between SR and ST as well as advancing on defining and operationalising these constructs encompasses the first step for standardising the research of these concepts, particularly across different domains of knowledge. At the same time, exposing the problems that emerge when defining and operationalising SR and ST across different domains and authors pushes up the standards of the research in these constructs by exerting pressure over the descriptive quality of experimental work in the field.

The second set of practical implications concern the fields of Psychology, Game Studies, and Learning Sciences. First, we worked in a methodological procedure for analysing commercial games in a comprehensive way; this is, accounting for the behaviour, representation, and cognitive demand of players when they are in a playing situation. This holistic understanding of games can be employed in future research about video games aiming for understanding the relationship between games and cognitive skills. Additionally, finding that two similar games cannot be directly compared as small the differences in game mechanics imply different behaviours and representation of the game by the players, derives to a question regarding skill transfer that must be raised in future game research. If skill transfer cannot be measured between two games, future research in video games aiming to test for skill transfer between games and other domains (including other video games) should develop new and more accurate measuring methods.

4.5. Recommendations for future research

4.5.1. Limitations

For the first phase of the study, we found three limitations: The extension of the sample, the deepening of domains, and the delimitation between commonalities between SR and ST.

The first limitation is due to the sampling method used. Although the method was chosen for giving a broad and efficient way of sampling texts from several databases from different disciplines, it lacks depth and reach. Future studies will benefit from deeper sampling methodology focusing on specific domains of knowledge with a more extensive data collection.

The second limitation is because we prioritise the exploration of the concepts in different domains. First of all, some domains focus more on researching these concepts (e.g., Psychology and Learning Sciences). Further studies will benefit from a deeper analysis of how the concepts are used in these fields to detect regularities and differences in the conceptualisations inside a specific domain for further comparison with other fields.

Regarding the second phase of the study, three limitations were found: The number of players analysed, the number of games analysed, and the type of games analysed.

The first limitation concerns the amount and level of players participating in the study. Although every player generated a large amount of quantitative information, the variance within the sample was not representative of different levels of expertise, particularly for the qualitative data. We suggest conducting further studies with a bigger sample in all different levels of expertise.

The second limitation concerns the use of only two MOBA games for the study. Although we decided to study two MOBAs that were very popular and highly identical, there are other MOBAs that can be researched and compared in regards to the use of cognitive skills, including MOBAs that are more dissimilar to the ones studied such as third person and first person ones.

The third limitation concerns the use of only one genre of games. There are many other genres of games that have not been studied concerning the cognitive demand, particularly for SR. The methodology developed for inquiring about commercial games and their relationship with SR could benefit from further testing with different game genres and gaming environments.

4.5.2. Future studies

The current advance made in the conceptual research allows for the use of content analysis of texts for identifying which constructs overlap, into which extent, and the way this occur. Therefore, we suggest further studies in the constructs of SR and ST for their deeper understanding, based on the current research findings and taking the proposed definitions as a starting point.

Additionally, the advances made with this research allow for the construction of a genealogical study aiming for tracing the origins and ramifications of the concepts of SR and ST, as well as the theoretical references quoted by authors studying these concepts. A genealogical study brings a historical account of how the concepts emerged, how they developed, and how they started overlapping.

Furthermore, we suggest a replication of this research using, first a different and more extensive sampling method for replicating or falsify the findings; second for deepening in particular domains of knowledge for understanding better how the concepts are used within fields.

Finally, we suggest the application of the conceptualisations of SR and ST to new pieces of research and to replication of research done before on SR and ST for further testing the adequacy of the definitions proposed as well as testing the fitness of other pieces of research which used a different operationalisation.

Regarding the second phase of the study, we suggest further research in MOBA games using a bigger and better-distributed sample to advance or falsify the findings of the present research. Additionally, we suggest conducting similar research with different MOBAs, first with the aim of testing the methodology used in the study to prove its fitness for studying MOBAs; and second, with the aim of testing how players of different MOBAs make use of SR or other cognitive abilities while playing games.

Furthermore, we suggest conducting research using the present methodology with different game genres to prove the fitness of the method for studying the use of cognitive skills, particularly SR, in other areas of game studies.

Additionally, we suggest the use of more up-to-date technology regarding eye-trackers sensitivity for evaluating the same type of actions using more fine-grained data and more AoIs. This technological improvement could increase our understanding of the way players search for information while playing and which type of information they seek.

Finally, we suggest conducting research involving professional eSports players to examine differences in SR and SM with ranked players and evaluate a wider range of expertise amongst the gaming population.

4.6. Considerations on the research of commercial video games

In this section, we summarise and reflect on the challenges and findings product of our research on commercial games. These findings, although not part of the main research, are a valuable contribution to future research in commercial games, particularly in the fields of Psychology and the Learning Sciences. We also offer a reflection on the way the challenges were tackled in the research with the aim to improve future research on commercial video games.

4.6.1. The nature of commercial video games and the challenges for their study

4.6.1.1. "Video game" is not one category

There are many types of video games genres, and there are different ways to categorise a game based on, for instance, its contents (e.g., horror game), the type of tasks performed by the player (e.g., puzzle game), the mechanics (e.g., point-and-click adventure), the population it is aimed to (e.g., family game), and even the platform the games run on (e.g., a Play Station game) (Papale, 2013). Therefore, playing *a* video game is not the same as playing *any* video game. It is necessary to take into account the type of game being studied and to restrict the variables as much as possible.

As an example, suppose someone studies how people react to horror games. It is not the same experience playing a *point-and-click horror adventure* in a *Play Station* compared to play a horror *First Person Shooter* on a desktop computer. More than that, there are different types of horror (i.e., gore horror, psychological horror, startle horror).

That video games can be fine-grained categorised has consequences for their research. Different approaches require different designs, samples, and have different degrees of generalisation. Consequently, general studies in games should broaden both the categories of games used in a study and the number of games considered for achieving a generalisation level. Conversely, it is possible to conduct a study about a single game title on a certain population.

For the present study, we overcame this obstacle by choosing a specific type of game (MOBAs). We chose LoL and DotA2 for being very popular and for being on the rise as an eSport. They are both structurally similar, using a system of lanes, towers, inhibitors, bases, and monsters. They both are played on an isometric grid, with the player looking at the map and minimap from above.

Although the games we studied are similar, we also found differences because the mechanics of the games are not entirely equal. As an example, players of both games answered differently concerning their hypotheses making and problem solving processes, indicating an emphasis on different points for which the game mechanics of each game depict different relevance.

4.6.1.2. A game is n-dimensional

The first challenge we found when studying video games is that they are not static. Although this is not only an issue for studying video games but for different phenomena, the non-static nature of video games makes them harder to study. The video game as a finished product might be considered as two or three-dimensional spaces (2 ½ for the isometric games). However, games also have an extra dimension: Time. Because of the time dimension, games are not finished products but a path that develops while the player travels it. Therefore, even if one person plays a game twice, the player is not playing the same game as their decisions make the game different. So when many players play a game, they are playing different paths of the same game (for an example of this issue see Canossa (2009), and Díaz and Tungjitcharoen (2015)). This case might not apply to Serious and Educational Games as they are often more structured, deterministic, and content oriented so that there is a minimum of strenuous stimuli related to the learning objective (Annetta, Minogue, Holmes, & Cheng, 2009).

Because of the n-dimensional nature of the games, we decided to evaluate the games (LoL and DotA2) from different perspectives using various types of data, triangulating the data so that we could understand both the game and the players. First of all, we decided to study how people look at the game as it serves as an indicator of cognitive activity (Poole & Ball, 2005). Eye-tracking methods also allow the creation of heat maps of the most relevant parts of the game for the player, as well as scan paths of procedural information seeking and strategy making. Additionally, we conducted semi-structured interviews for better understanding how players represent the game, their intentions, motivations, and strategies behind their play; thus, allowing to discern their cognitive processes in a more comprehensive way.

This holistic approach is useful for accounting for many dimensions of the game in a processual way. Nevertheless, it has the downside of producing a big amount of data that is time-consuming to process and analyse.

4.6.1.3. The game experience is not static

Another challenge found when studying video games is that time passes, not only in the video game but also in everyday life. Although this issue does not only attain game studies, it is important to account for it as the media technologies can change quickly (See Bartle (2015) for a reference to this issue). There are three main issues to account for when conducting research in games: (1) If the researcher is studying development, it is not possible to prevent that the participants play or inform themselves about the game in between evaluation sessions. Conversely, in some situations, the evaluator would like players to play more of a specific game at home, but instead, players do not play it or play other types of games (see for example Steinkuehler & Duncan, (2008)). (2) If the researcher is studying games with a social aspect, like a Massive Multiplayer Online Role Playing Game (MMORPG), the social interactions within a group might change from one session to the next, thus changing the way people play the game (for more information see Steinkuehler (2008)). (3) If the researcher is investigating a game with periodic updates, they might find themselves evaluating a different version of the game from the first to the last day of evaluation. Furthermore, sometimes big updates can happen overnight.

In the present research, we chose to embrace the chaotic nature of the games (development of expertise, social interaction, and periodic updates) and documenting it instead of trying to prevent it. Moreover, to know the full potential of the expert players, we allowed them to take their playing hardware to the laboratory (i.e., gamer mouse) and to shape the GUI as they needed. The rationale for this was that, although fixed rules create a situation easier to evaluate, these rules can limit the potential of the players. As examples, we observed participants making the GUI very small, so they had a bigger overview of the game; we also found participants shifting the LoL minimap to the left side of the screen as they were not used to play with it on the right side.

As the game also contained periodic updates, we informally inquired expert participants about the extent of updates of the game and if they interfered with their normal game play. Although there were many updates in the LoL (v 3.14 – v 4.20) and DotA2 (v 6.83b – v 6.84) games in the eight-month period of data collection, no player reported any influence of the updates on their gameplay.

4.6.1.4. The game is also the Metagame

Playing a game, and especially learning a game is not only about sitting playing for hours. Societies and networks of knowledge are built both inside and outside games to explain game mechanics, guide players, improve strategies (see Steinkuehler (2008) for example). A player does not just learn to play. Consequent to the nature of the game they look for knowledge tools on the internet, their friends, or people inside the game, aiming to boost their proficiency of the game beyond the single playing practice (Steinkuehler, 2008), just as Vygotsky describes the Zone of Proximal Development (Vygotsky, 1979).

4.6.1.5. Gaming is an experience and a process

Games differ in complexity, content, depth, mechanics, and more; but there are two aspects of a game that are always present (1) the time dimension and (2) the individual (player-subjective) dimension. The first refers to the fact that a game is not static, but unfolds itself when played. The second refers to the representations that individuals have concerning a specific game. Player representations of a game can be different from one game session to another, and the representations generated by a game can be different from one person to another (Bartle, 2015). Because of these two aspects, games cannot be treated as a single stimulus. Asking about how people play, their motivations, and representation of the game, create more understanding about what players learn from the game and which skills they use and develop than static data such as the number of character deaths or the time required for game completion (Lieberoth & Roepstorff, 2015).

In the present research we aimed for a game environment familiar to the players, so they did not feel under pressure, therefore fostering the regular gaming experience for that particular player. To evaluate the process, we used every session as an evaluation in itself and followed with qualitative interviews.

Additionally, we used a mixed etic and emic perspective (Pike, 1967) by learning and familiarising with the games and the culture around them. The mixed perspective allowed us to talk freely with the participants, understanding their jargon, affordances, and limitations. In other words, confer with the participants in a player-player relationship and not in an evaluator-evaluated relationship.

4.6.1.6. Gaming is a non-linear

Researching learning and development of cognitive skills in commercial games is confusing due to the amount of variables and dimensions interacting in time. For example, a person can learn to play a game by playing it, watching videos on the internet, take shortcuts, cheat, get information from friends, all of this while playing other games in parallel. Playing a commercial game is not about sitting in front of the computer defeating bosses. Depending on the type of game and the type of player, some people will choose to investigate how the game world works, others would rather talk in the chat, and others would go to the libraries inside the game to learn the story of a forgotten empire (Bartle, 2013). All players are learning, developing or exploiting different skills in an interrelation of various components that change over time and which knowledge is distributed amongst gameplay, videos, walkthroughs, and memes (Gee, 2007a).

The challenge is that research aims for finding patterns inside the gameplay such as how people use the minimap, how people gather information to make hypotheses and strategies, or how they communicate these strategies. The game researcher aims to understand the interaction of game elements and people, how the player solves problems and overcomes difficult situations; in other words, they aim to investigate both the (game and cognitive) elements and their relations, and the processes that emerge from such interactions.

The present research used a holistic approach to account for the cognitive processes of the participants as well as commonalities and differences, both from an individual and a group perspective. The goal with this is to detect how players make use of SR during the gameplay, discovering systematicity in patterns of actions that can be generalised, and emergent patterns product of the interaction between the player and the game.

4.6.2. Research design and methodological problems

4.6.2.1. Categorising video games for research

The first issue when researching commercial video games is how to categorise them. The standard scientific method (Godfrey-Smith, 2003) consists of isolating a variable to determine different factors, for instance, the development under different conditions, if there is an intervention effect, if two variables correlate, amongst others. Then generalise and create models of understanding for the phenomenon (Godfrey-Smith, 2003). Therefore, the first challenge when researching video games is to categorise them.

Categorising video games is a problem because of the different ways there are to do it and how some game genres overlap. Retaking the horror games example, we can start from the idea that there are different genres of horror. Narrowing down the categories is possible to find out that there are point and click, first person, third person and platform horror games (amongst others). Narrowing down more, it is not the same experience to play with a keyboard, an analogue control, or a Virtual Reality device with movement sensors.

If we follow the basic idea of generalisation from the standard scientific method, to study video games we would need to: (a) Have a considerable sample of players for every game in that genre, so that generalisations can be made despite the game style or platform. (b) Narrow the genre variable taking into account other facts such as game style and platform, thus lowering the generalisation to only these variables. (c) Focus on one game, making it a case study or a title study, thus disregarding generalisation but broadening the understanding of a particular experience.

4.6.2.2. N-dimensions and dimension reduction

Reducing game dimension from N to three or four dimensions for evaluation is difficult. If we research the game as a process, we will get several hours of video recordings of different games, along with some interviews. If we are making an MMORPG ethnography (Martin et al., 2011), we might end up with several log files of chats and recordings of conversations via the internet.

Reducing the dimensions of a game for studying it is challenging and does not necessarily mean that the phenomenon studied is clearer. Also, the dimensions reduced to depend on the research question. If the research question is operationalised so that it can be directly answered by the outcome of the game process (for instance, the relation between the age of players and the scores achieved in a game) then reducing the dimensions to these two is the best solution. Nevertheless, a more complex research question (for instance, if playing a game enhances spatial reasoning) would require a comprehensive task analysis of the game and the reasoning processes of the participants (Díaz, 2012; Newell, 1966; Newell & Simon, 1972).

4.6.2.3. Normalised vs. Adaptable research

Studying a video game as a static event that took place once in the experience of several people under controlled circumstances is different to study games as a process open to any changes. Different ways to tackle this problem have implications for the quality, validity, and generalisation of the results. Players change tactics depending on their mood, the game can change from one day to the next because of major server updates, the internet can fail during an evaluation session, and players can explore other aspects of the game outside the evaluation session.

The complexity of games does not mean they cannot be simplified to an outcome or a static process, but this simplification has to be in line with the research question. Also, the sample size plays a major role in accounting for variability and generalisability of the results. When studying a process the variance of outcomes at different stages of the process can be so big that data can be lost in the variance. As an example, think about which moves can a chess player make in their third turn; how many players will play a specific move and why; or how is this action influenced by the previous moves (system's past) or the player strategy (system's forecast). We recognise different ways to overcome this problem. (1) Making a detailed game analysis of the interactions between the game parts (e.g., mechanics, choices, and decision trees) and the possible outcomes of each decision, and then confront them with expert reasoning (Newell, 1966; Newell & Simon, 1972). (2) Having bigger samples so that trends can be drawn from the variability. (3) Understand how players learn the game, construct their problem space (Newell & Simon, 1972), and adapt to new situations in the game progression so processes can be inferred from the gaming experience. (4) Creating a systemic simulation accounting for different procedures and interaction between the game parts and the players, allowing for elucidating different behaviours, including emergent ones.

4.6.2.4. Considerations about the inclusion of the metagame in research

Game researchers have documented how the metagame has value not only for learning a game but also for improving the performance of the player within the game (Steinkuehler, 2008, 2010).

When evaluating a player in a controlled setting, the usual decision is to deny the player access to the metagame (e.g., forbidding reading or watching videos about the game). Other types of research such as Naturalistic research or Ethnography would allow players to explore and talk about it as they want. The first perspective gives the researcher more control over the variables as they can choose what information to give to the player and focus on certain variables. Nevertheless, this can cap the learning of the player and the performance in the game. In contrast, giving players access to extra information about the game might encourage the deployment of new strategies and increase the performance of the player, but lowering the control of the research over the main game variables.

4.6.2.5. Process evaluation or outcome evaluation

Games are processes (Bartle, 2015), but they are often evaluated as outcomes independent of the circumstances surrounding them (e.g., scoreboards, performance ratios, and psychological tests) (Alba-Marrugo, 2013). Although evaluating the game as an outcome has the advantage of being direct and relatively quick, it is also very limited. Evaluating a game as a process implies more work from the research team, and the trade-off in generalisation might not be good; however, it is a suitable methodology for understanding the game, the player interaction, and the mechanics game-player.

4.6.2.6. The correct time for questions

The form of evaluation is also a trade-off when it comes to research commercial games. Direct evaluation often leads to a disruption of the game experience or extra effort from the player that changes the way they regularly play (van Gog, Paas, van Merriënboer, & Witte, 2005). As an example, Think Aloud Protocols require training the player for correctly performing the thinking aloud task; they also increase the cognitive load in the game, creating a negative interaction with the task performance (van den Haak, De Jong, & Schellens, 2010; van Gog et al., 2005).

Evaluating the player after finishing a play session with self-reports can influence answers because of bias, confabulations, or made up answers that might not be the correct indication of the processes happening when playing (Delroy & Simine, 2007; Efstratios, 2010). From our experience, the use of Cued Retrospective Reports using eye-tracking videos as elicitors brings a good balance to the evaluation of the gameplay. It does not interfere with the gameplay, cues the player to the place they are looking at a certain point increasing the accuracy of the report and allows the researcher to detect confabulation of memories in the reports (Dado, 2015).

4.6.2.7. Skill transfer revisited

Although this problem is not only part of game studies and has been highly documented in other disciplines (Scardamalia & Bereiter, 2006), it remains a central challenge when studying commercial games. In fact, it is sometimes an issue for Serious Games (McClarty et al., 2012).

When researching games it is often assumed that a player acquires certain sets of skills when playing, which can be comparatively evaluated using analogue tasks or standardised tests (Alba-Marrugo, 2013; Glass et al., 2013; Irons et al., 2011). Nevertheless, even if the skills developed are similar, the skill transfer can be far enough that the player cannot solve the evaluation task according to the skills developed by playing (Baldwin & Ford, 1988; Barnett & Ceci, 2002; Boot et al., 2008). As there is not yet a normalised way to measure the amount of skill transfer (particularly in the field of video games), there is a possibility that researchers measure a different construct to what is developed within the gaming environment.

Based on our experience with the current study, evaluation of skill transfer in games is problematic. For instance, although we chose two games that were structurally similar in genre, design, gameplay, perspective, and platform; it turned out that the two games are played differently and involve different cognitive abilities from one another. This difference in gameplay is the result of small differences regarding the use of spaces, money, and items, which creates a breach in the comparative evaluation of players.

It is possible to implement research designs aiming to compare the performance and development of novice players and expert players, but this type of design has the disadvantage of being game-centred. That is to say; skill transfer is not being evaluated, so it is not possible to derive conclusions about performance in other games, formal settings, or real life situations. If the goal of the research is to evaluate skill transfer, we suggest exploring, developing, and using other methods that could account for close and far transfer of skills (Boot et al., 2008) (see Barnett & Ceci (2002) & van Lehn (1996) for further assessing this idea).

4.6.2.8. Researching non-linear systems

Games are systems: A complex interaction between game mechanics, player-game interaction, metagame, game updates, play styles, and social interactions. All systems can be studied in different ways, from simplified versions of the system to models depicting the interaction of the structures (Gharajedaghi, 2011). Using a systemic approach for studying a phenomenon also allows for uncovering and highlighting emergent behaviours from the system; behaviours that cannot be observed by studying independent characteristics of the phenomenon.

Based on our experience studying games as processes, we suggest to document all changes and to have an emic perspective when studying the game (meaning that the researcher should play and be familiar with the game, the game scene, and the jargon before studying it). It is also important to establish an open communication with the participants regarding changes in the game, metagame activities, perceptions they have regarding the game and the environment of study. Accounting for these elements allows the researcher to recognise interactions between the system's (game-player) elements, as well as emergent phenomena product of their interaction.

4.6.2.9. Reliability: Balancing significant sampling and amount of data

The last problem discussed here belongs to the field of statistical reliability. Studies with large samples are required to make generalisations and to acquire statistic significance. Although studies with large samples are required to account for generalisability of trends and patterns in the behaviour of players, when trying to understand the game as a process the amount of data preparation and analysis becomes a challenge. Understanding a game as a process requires of different levels of analysis, which is costly regarding the sample, space, and data processing time. Therefore, the bigger the sample required for higher reliability, the more resources needed and the more difficult becomes to process and analyse the data (Lieberoth & Roepstorff, 2015).

As researchers we aim for a high validity and reliability in our studies; therefore, we aim for a significant sample allowing for generalisation. Nevertheless, we should not forget that other types of research also help to explore phenomena in depth; such is the concern of case studies. These types of studies should be welcomed as building stones for theories within psychological game studies.

Finally, systematically reporting on research procedures will help the community of game studies (particularly in the fields of Psychology and Learning Sciences) to develop better tools and methodologies for understanding what happens in the gamer's mind when playing, the relationship game-player, as well as the development of better games.

References

Abdulkarim, R., & Al Jadiri, A. (2012). The effect of cooperative learning group division based on multiple intelligences theory and previous achievement on scientific thinking skills development of ninth grade students in Oman. *European Journal of Social Sciences*, *27*(4), 553–569.

Abdullah, S., & Shariff, A. (2008). The effects of inquiry-based computer simulation with cooperative learning on scientific thinking and conceptual understanding of gas laws. *Eurasia Journal of Mathematics, Science and Technology Education*, *4*(4), 387–398.

Acar, O., & Patton, B. (2012). Argumentation and formal reasoning skills in an argumentation-based guided inquiry course. *Procedia-Social and Behavioral Sciences*, *46*, 4756–4760.

Adey, P., & Csapó, B. (2013). Chapter 1. Developing and assessing scientific reasoning. In B. Csapó & G. Szabó (Eds.), *Framework for Diagnostic Assessment of Science* (pp. 17–53).

Aiello, M. (2002). *Spatial reasoning: Theory and practice*. Universiteit van Amsterdam, http://www.cs.rug.nl/~aiellom/publications/aielloPHDThesis.pdf.

Akkerman, S., Admiraal, W., Brekelmans, M., & Oost, H. (2008). Auditing quality of research in social sciences. *Quality & Quantity*, *42*(2), 257–274. https://doi.org/10.1007/s11135-006-9044-4

Alba-Marrugo, M. (2013). *Cambios en el componente de la flexibilidad cognitiva de la función ejecutiva en niños entre 6 y 8 años, luego de exposición a videojuegos de* *gestión de tiempo* [Changes in the cognitive flexibility component of executive functions in children between 6 and 8 years old after being exposed to time management video games] (Master thesis). Universidad de Antioquia, Colombia.

Alonso, G. (2013). Revisión del concepto de desarrollo local desde una perspectiva territorial [Revision of the concept of 'local development' from a territory perspective]. *Revista Líder*, *23*, 9–28.

Anderson, C. (2003). Violent video games: Myths, facts, and unanswered questions. *Psychological Science Agenda*, *16*(5).

Anderson, C., & Dill, K. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology*, *78*(4), 772–290. https://doi.org/10.1037//O022-3514.78.4.772

Aneta, K., & Jerzy, P. (2013). Abductive and deductive approach in learning from examples method for technological decisions making. *Procedia Engineering*, *57*, 583–588. https://doi.org/10.1016/j.proeng.2013.04.074

Annetta, L., Minogue, J., Holmes, S., & Cheng, M.-T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education*, *53*, 74–85.

Ato, M., López, J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología [A classification system for research designs in Psychology]. *Anales de Psicología*, *23*(3), 1038–1059. https://doi.org/10.6018/analesps.29.3.178511

Azarpira, N., Amini, M., Kojuri, J., Pasalar, P., Soleimani, M., Kahni, S., … Lankarani, K. (2012). Assessment of scientific thinking in basic science in the Iranian second national Olympiad. *BCM Research Notes*, *5*(1), 1–7.

Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development*, *2*(3), 201–221. https://doi.org/10.1111/j.1467-9507.1993.tb00014.x

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, *41*(1), 63–105. https://doi.org/10.1111/j.1744-6570.1988.tb00632.x

Baniqued, P., Lee, H., Voss, M., Basak, C., Cosman, J., DeSouza, S., … Kramer, A. (2013). Selling points: What cognitive abilities are tapped by casual video games? *Acta Psychologica*, *142*(1), 74–86.

Barnett, S., & Ceci, S. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. https://doi.org/10.1037//0033-2909.128.4.612

Barnett, S., & Ceci, S. (2005). The role of transferable knowledge in intelligence. In R. Sternberg & J. Pretz (Eds.), *Cognition and Intelligence. Identifying the Mechanisms of the Mind*. New York, NY, USA: Cambridge University Press.

Bartle, R. (2013, September 23). Hearts, clubs, diamonds, spades: players who suit MUDs. Retrieved from http://www.mud.co.uk/richard/hcds.htm

Bartle, R. (2015). Special issue issues. *Psychology of Popular Media Culture*, *4*(4), 259–276. https://doi.org/10.1037/ppm0000078

Bavelier, D., Green, S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: learning to learn and action video games. *The Annual Review of Neuroscience, 35*, 931–416. https://doi.org/10.1146/annurev-neuro-060909–152832

Beatty, E., & Thompson, V. (2012). Effects of perspective and belief on analytic reasoning in a scientific reasoning task. *Thinking and Reasoning*, *18*(4), 441–460. https://doi.org/10.1080/13546783.2012.687892

Bioware. (2009). *Dragon Age* [Video game].

Blair, G., & Goodson, M. (1939). Development of scientific thinking through general science. *The School Review*, *47*(9).

Blumberg, F., Altschuler, E., Almonte, D., & Mileaf, M. (2013). The impact of recreational video game play on children's and adolescents' cognition. *New Directions for Child and Adolescent Development*, *139*, 41–50. https://doi.org/10.1002/cad.20030

Bolzer, M., Strijbos, J.-W., & Fischer, F. (2015). Inferring mindful cognitive-processing of peer-feedback via eye-tracking: role of feedback-characteristics, fixation-durations and transitions. *Journal of Computer Assisted Learning*, *31*, 422–434. https://doi.org/10.1111/jcal.12091

Bonny, J., Castaneda, L., & Swanson, T. (2016). Using an international gaming tournament to study individual differences in MOBA expertise and cognitive skills. In *Proceedings of CHI 2016* (pp. 3473–3484). San José, California, USA. https://doi.org/http://dx.doi.org/10.1145/2858036.2858190

Boot, W., Kramer, A., Simons, D., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica*, *129*(3), 387–398. https://doi.org/10.1016/j.actpsy.2008.09.005

Borecki, L., Tolstych, K., & Pokorski, M. (2013). Computer games and fine motor skills. Advances in experimental medicine and biology, 755, 343.

Bourdieu, P. (1991). The peculiar history of scientific reason. *Sociological Forum*, *6*(1), 3–26.

Brigandt, I. (2010). Scientific reasoning is material inference: Combining confirmation, discovery, and explanation. *International Studies in Philosophy of Science*, *24*(1), 31–43. https://doi.org/10.1080/02698590903467101

Brown, N., Furtak, E., Timms, M., Nagashima, S., & Wilson, M. (2010). The evidence-based reasoning framework: Assessing scientific reasoning. *Educational Assessment*, *15*(3), 123–141. https://doi.org/10.1080/10627197.2010.530551

Burak, A., Keylor, E., & Sweeney, T. (2007). Pacemaker: A video game to teach peace. *Intelligent Technologies for Interactive Entertainment Proceedings*, *3814*, 307–310.

Callahan, J. (2014). Writing literature reviews: A reprise and update. *Human Resource Development Review*, *13*(3), 271–275. https://doi.org/10.1177/1534484314536705

Canossa, A. (2009). *Play-Persona: Modelling player behaviour in computer games*. Danmarks Design Skole. Retrieved from http://www.dcdr.dk/dk/materiale/dokumenter/phd-afhandling-alessandro-canossa.pdf

Car, A., & Frank, A. (1994). General principles of hierarchical spatial reasoning: The case of wayfinding. In *Proceedings of the conference on Spatial Data Handling* (pp. 646–664). Edinburg, Scotland. Retrieved from ftp://ftp.geoinfo.tuwien.ac.at/car/acSDH94.pdf

Chi, M. T. H., Glaser, R., & Farr, M. (Eds.). (1988). *The nature of expertise*. New Jersey, USA.: Lawrence Erlbawm Associates, Inc.

Ciavarro, C., Dobson, M., & Goodman, D. (2008). Implicit learning as a design strategy for learning games: Alert Hockey. Computers in Human Behavior, 24(6), 2862-2872

Cocking, R., Mestre, J., & Brown, A. (2000). New developments in the science of learning: Using research to help students learn science and mathematics. *Journal of Applied Developmental Psychology*, *21*(1), 1–11.

Coletta, V., Jeffrey, P., & Steinert, J. (2011). FCI normalized gain, scientific reasoning ability, thinking in physics, and gender effects. In *Proceedings of the 2011 Physics Education Research Conference* (Vol. 1413, pp. 23–26). Nebraska, USA. https://doi.org/10.1063/1.3679984

Collings, J. (1985). Scientific thinking through the development of formal operations: training in the cognitive restructuring aspect of field. *Research in Science and Technological Education*, *3*(2). https://doi.org/10.1080/0263514850030107a

Colman, A. (2003). Depth of strategic reasoning in games. *Trends in Cognitive Sciences*, *7*(1).

Connolly, T., Boyle, E., MacArthur, E., Hainey, T., & Boyle, J. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers and Education*, *59*, 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

Cortázar, J. (1969). Instructions on how to climb a staircase. In *Cronopios and Famas* (pp. 21–22). CPI group.

Creswell, J. (2015). *A concise introduction to mixed methods research*. USA: Sage Publications, Inc.

Croker, S., & Buchanan, H. (2011). Scientific reasoning in a real-world context: the effect of prior belief and outcome on children's hypothesis-testing strategies. *British Journal of Developmental Psychology*, *29*(3), 409–424. https://doi.org/10.1348/026151010X496906

Crowley, K., Callanan, M., Jipson, J., Galco, J., Topping, K., & Shrager, J. (2001). Shared scientific thinking in everyday parent-child activity. *Science Education*, *85*(6), 712–732. https://doi.org/10.1002/sce.1035

Dado, M. (2015). *Learning in technologies: Situation awareness in Massively Multiplayer Online Games* (Master thesis). Ludwig-Maximilians-University Muenchen, Munich, Germany.

de Castell, S., & Jenson, J. (2007). Digital games for education: When meanings play. *Intermédialités: Histoire et Théorie Des Arts, Des Lettres et Des Techniques*, (9), 113. https://doi.org/10.7202/1005533ar

Dejonckheere, P., van de Keere, K., & Mestdagh, N. (2010). Training the scientific thinking circle in pre- and primary school. *The Journal of Education Research*, *103*, 1–16. https://doi.org/10.1080/00220670903228595

Delroy, P., & Simine, V. (2007). Chapter 13. The self-report method. In *Handbook of research methods in personality psychology* (pp. 224–239). New York, USA: Guildford press.

Devís-Devís, J., Peiró-Velert, C., Beltrán-Carrillo, V., & Tomás, J. (2009). Screen media time usage of 12-16 year-old Spanish school adolescents: Effects of personal and socioeconomic factors, season and type of day. *Journal of Adolescence*, *32*(2), 213–231.

Díaz, C. M. (2012). *Development of a methodological tool for analysing the reasoning demand of video games in an agile and structural manner* (Master thesis). University of Copenhagen, Copenhagen, Denmark.

Díaz, C. M. (2015, October). MOBA-ET-Repository [Internet Repository]. Retrieved from https://github.com/cmauricio/MOBA-ET-R

Díaz, C. M. (2016, July 14). ConceptReview-Repository [Internet Repository]. Retrieved from https://github.com/cmauricio/ConceptReview

Díaz, C. M., & Tungjitcharoen, W. (2015). Art video games ritual communication of feelings in the digital era. *Games and Culture*, *10*(1), 3–34. https://doi.org/10.1177/1555412014557543

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153–170. https://doi.org/10.1016/j.learninstruc.2012.05.003

Dogusoy-Taylana, B., & Cagiltay, K. (2014). Cognitive analysis of experts' and novices' concept mapping processes: An eye tracking study. *Computers in Human Behavior*, *36*, 82–93. https://doi.org/010.1016/j.chb.2014.03.0360747-5632

Downing, E. (1928). The elements and safeguards of scientific thinking. *The Scientific Monthly*, *26*(3), 231–243.

Duchowski, A. (2007). Eye tracking methodology: theory and practice (2nd ed.). Great Britain: Springer-Verlag London Limited.

Dunbar, K. (2000). How scientists think in the real world: Implications for science education. *Journal of Applied Developmental Psychology*, *21*(1), 49–58.

Dunbar, K. (2001). Chapter 5. What scientific thinking reveals about the nature of cognition. In K. Crowley, C. Schunn, & T. Okada (Eds.), *Designing for Science: Implications from Everyday, Classroom, and Professional Settings* (pp. 115–140). Psychology Press.

Dunbar, K., & Klahr, D. (2012). Chapter 35. Scientific thinking and reasoning. In K. Holyoak & R. Morrison (Eds.), *Oxford Handbook of Thinking and Reasoning* (2012 ed.). New York, USA: Oxford University Press.

Durkin, K., & Barber, B. (2002). Not so doomed: Computer game play and positive adolescent development. *Journal of Applied Developmental Psychology*, *23*(4), 373–392.

Durso, F., & Gronlund, S. (1999). Chapter 10: Situation awareness. In F. Durso, R. Nickerson, R. Schvaneveldt, S. Dumais, D. Lindays, & M. Chi (Eds.), *Handbook of applied cognition* (pp. 289–314). Wiley.

Dye, M. W. G., Green, C. S., & Bavelier, D. (2009). The development of attention skills in action video game players. *Neuropsychologia*, *47*(8–9), 1780–1789. https://doi.org/10.1016/j.neuropsychologia.2009.02.002

Efstratios, T. (2010). *Let the gamers do the talking: A comparative study of two usability testing methods for video games* (Master thesis). University College of London, UK.

Elo, A. (2008). *The rating of chess players, past and present* (2nd ed.). Arco Publications.

Endsley, M. (1990). Predictive utility of an objective measure of situation awareness. In *Proceedings of the Human Factors Society*, (Vol. 34, pp. 41–45).

Engelhardt, C., Bartholow, B., Kerr, G., & Bushman, B. (2011). This is your brain on violent video games: Neural desensitization to violence predicts increased aggression following violent video game exposure. *Journal of Experimental Social Psychology*, *47*, 1033–1036. https://doi.org/10.1016/j.jesp.2011.03.027

Entertainment Software Association ESA. (2016). *The 2016 essential facts about the computer and video game industry* (Sales, Demographics, and Usage Data). Retrieved from http://essentialfacts.theesa.com/Essential-Facts-2016.pdf

Etkina, E., Karelina, A., & Villasenor, M. (2007). Studying transfer of scientific reasoning abilities. In *2006 Physics Education Research Conference* (pp. 81–84). New York, USA: AIP Publishing. https://doi.org/http://dx.doi.org/10.1063/1.2508696

Eysenck, M., & Keane, M. (2003). *Cognitive Psychology A student's handbook* (4th ed.). USA: Psychology Press.

Facer, K., Joiner, R., Stanton, D., Reid, J., Hull, R., & Kirk, D. (2004). Savannah: mobile gaming and learning? *Journal of Computer Assisted Learning*, *20*, 399–409.

Faulkner, D., Joiner, R., Littleton, K., Miell, D., & Thompson, L. (2000). The mediating effect of task presentation on collaboration and children's acquisition of scientific reasoning. *European Journal of Psychology of Education*, *15*(4), 417–430. https://doi.org/10.1007/BF03172985

Fender, J., & Crowley, K. (2007). How parent explanation changes what children learn from everyday scientific thinking. *Journal of Applied Developmental Psychology*, *28*(3), 189–210. https://doi.org/10.1016/j.appdev.2007.02.007

Ferguson, C. (2007). Evidence for publication bias in video game violence effects literature: A meta-analytic review. *Aggression and Violent Behavior*, *12*(4), 470–482. https://doi.org/10.1016/j.avb.2007.01.001

Ferguson, C., Coulson, M., & Barnett, J. (2011). A meta-analysis of pathological gaming prevalence and comorbidity with mental health, academic and social problems. *Journal of Psychiatric Research*, *45*(12), 1573–1578. https://doi.org/10.1016/j.jpsychires.2011.09.005

Fernie, D., & DeVries, R. (1990). Young children's reasoning in games of nonsocial and social logic:'Tic Tac Toe' and a 'Guessing Game'. *Early Childhood Research Quarterly*, *5*(4).

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., … Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, *2*(3), 28–45. https://doi.org/http://dx.doi.org/10.14786/flr.v2i2.96

Fischer, S., Hickey, D., Pellegrino, J., & Law, D. (1994). Strategic processing in dynamic spatial reasoning tasks. *Learning and Individual Differences*, *6*(1), 65–105.

Fraps, & Beepa. (2013). Fraps (Version 3.5.99) [Windows]. Retrieved from http://www.fraps.com/faq.php

Frasca, G. (2001). Rethinking agency and immersion: Videogames as a means of consciousness-raising. Digital Creativity, 12(2), 167–174.

Fudenberg, D., & Levine, D. K. (1998). The theory of learning in games (Vol. 2). MIT press

Fullerton, T. (2008). *Game Design Workshop*. USA: Elsevier.

Furley, P., & Wood, G. (2016). Working memory, attentional control, and expertise in sports: A review of current literature and directions for future research. *Journal of Applied Research in Memory and Cognition*, 1–11. https://doi.org/10.1016/j.jarmac.2016.05.001

Gagnon, D. (1985). Videogames and spatial skills: An exploratory study. Educational Communication and Technology, 33(4), 263–275.

Galarneau, L. L. (2005). Authentic learning experiences through play: Games, simulations and the construction of knowledge. Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=810065

Gallardo-Echenique, E., Marqués-Molías, L., Bullen, M., & Strijbos, J.-W. (2015). Let’s talk about digital learners in the digital era. *International Review of Research in Open and Distributed Learning*, *16*(3), 156–187.

Garnham, A., & Oakhill, J. (1994). *Thinking and Reasoning*. USA: Basil Blackwell Inc.

Gee, J. P. (2005a). *Learning about learning from a video game: Rise of nations*. Essay. Retrieved from http://www.academiccolab.org/resources/documents/RON-paper.rev.pdf

Gee, J. P. (2005b). *Video games, mind, and learning*. Essay, Wisconsin, United States of America. Retrieved from http://academiccolab.org/resources/documents/IDMA_Paper.pdf

Gee, J. P. (2005c). *Why video games are good for your soul*. Australia: Common Ground Publishing Pty Ltd.

Gee, J. P. (2006). Are video games good for learning? Presented at the 13th national conference, Wisconsin, USA, Wisconsin, USA. Retrieved from http://cmslive.curriculum.edu.au/verve/_resources/Gee_Paper.pdf

Gee, J. P. (2007a). *What video games have to teach us about learning and literacy*. New York, USA: Library of Congress.

Gee, J. P. (2007b). *Why are video games good for learning*. Essay. Retrieved from http://www.academiccolab.org/resources/documents/MacArthur.pdf

Gentile, D. A. (2011). The multiple dimensions of video game effects. Child development perspectives, 5(2), 75-81.

Gharajedaghi, J. (2011). *Systems thinking* (3rd ed.). USA: Elsevier. Retrieved from http://www.sciencedirect.com/science/book/9780123859150

Gilbert, S., & Burgess, P. (2008). Executive function. *Current Biology*, *18*(3), R110–R114. https://doi.org/10.1016/j.cub.2007.12.014

Glass, B., Maddox, T., & Love, B. (2013). Real-Time Strategy Game training: Emergence of a cognitive flexibility trait. *PLOS One*, *8*(8), e70350. https://doi.org/doi:10.1371/journal.pone.0070350

Gleason, M., & Schauble, L. (1999). Parents' assistance of their children's scientific reasoning. *Cognition and Instruction*, *17*(4).

Godfrey-Smith, P. (2003). *Theory and Reality. An introduction to the philosophy of science*. USA: The University of Chicago Press.

Goldberg, J., & Kotval, X. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, *24*(6), 631–645.

Goldberg, J., Stimson, M., Lewenstein, M., Scott, N., & Wichansky, A. (2002). Eye tracking in web search tasks: Design implications. (pp. 51–58). New York, NY, USA: ACM Press.

Gollwitzer, M., & Melzer, A. (2012). Macbeth and the Joystick: Evidence for moral cleansing after playing a violent video game. *Journal of Experimental Social Psychology*, (48).

Gower, B. (1997). Henri Poincare and Bruno de Finetti: Conventions and scientific reasoning. *Studies in History and Philosophy of Sciences*, *28*(4), 657–679.

Green, C. S., & Bavelier, D. (2012). Learning, attentional control, and action video games. Current Biology, 22, R197–R206. https://doi.org/10.1016/j.cub.2012.02.012

Green, S., Pouget, A., & Bavelier, D. (2010). Improved probabilistic inference as a general learning mechanism with action video games. Current Biology, 20, 1573–1579. https://doi.org/10.1016/j.cub.2010.07.040

Greenfield, P., Brannon, C., & Lohr, D. (1994). Two-Dimensional representation of movement through three-dimensional space: The role of video game expertise. Journal of Applied Developmental Psychology, 15, 87–103.

Griffiths, M. (2002). The educational benefits of videogames. Education and Health, 20(3), 47–51. Retrieved from http://dh101.humanities.ucla.edu/DH101Fall12Lab4/archive/files/6070f0882cc89baa5a12992aaea56028.pdf

Hamari, J., & Sjöblom, M. (2015). What is eSports and why do people watch it? *Social Science Research Network*, 26. Retrieved from http://ssrn.com/abstract=2686182

Harre, R. (2004). Chapter 3. Properties and Images. In D. Rothbart (Ed.), *Modeling: Gateway to the Unknown* (Vol. 1, pp. 29–45). Elsevier Inc.

Helm, B. (2005). Educational games crank up the fun. Retrieved 1 January 2013, from http://www.businessweek.com/stories/2005-08-22/educational-games-crank-up-the-fun

Hofferth, S., & Sandberg, J. (2001). How American children spend their time. *Journal of Marriage and Family*, *62*(2), 295–308.

Holmes, N., & Bonn, D. (2013). Doing science or doing a lab? Engaging students with scientific reasoning during physics lab experiments. In *Proceedings of the 2013 Physics Education Research Conference*. Portland USA. https://doi.org/10.1119/perc.2013.pr.034

Holyoak, K., & Morrison, M. (2005). *The Cambridge handbook of Thinking and Reasoning*. USA: Cambridge University Press.

Hong, J.-C., & Liu, M.-C. (2003). A study on thinking strategy between experts and novices of computer games. *Computers in Human Behavior, 19*(2), 245–258.

Hopkins, J. (1996). Psychoanalytic and scientific reasoning. *British Journal of Psychotherapy, 13*(1), 86–105.

Hsieh, W.-M., Cui, L., & Sharma, P. (2008). Facilitating critical and scientific thinking using group-based problem solving activities in a large undergraduate course. In *Selected*

Research, Development, and Papers on the Practice of Educational Communications and Technology (pp. 124–133). Orlando, USA.: The Research and Theory Division.

Huizinga, J. (1950). *Homo Ludens. A study of the play element in culture* (2014 ed.). Connecticut, USA: Martino Publishing.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. Psychology Press.

Irons, J., Remington, R., & McLean, J. (2011). Not so fast: Rethinking the effects of action video games on attentional capacity. *Australian Journal of Psychology*, *63*(4), 224–231. https://doi.org/10.1111/j.1742-9536.2011.00001.x

Jirout, J., & Klahr, D. (2012). Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental Review*, *32*, 125–160. https://doi.org/10.1016/j.dr.2012.04.002

Johnson, T., & Krems, J. (2001). Use of current explanations in multicausal abductive reasoning. *Cognitive Science*, *25*, 903–939.

Johnson-Laird, P. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge, United Kingdom: Cambridge University Press.

Johnson-Laird, P., & Byrne, R. (1993). Mental models or formal rules? *Behavioral and Brain Sciences*, *16*(2), 368–380. https://doi.org/10.1017/S0140525X0003065X

Jonassen, D. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development*, *45*(1), 65–94. https://doi.org/10.1007/BF02299613

Jonassen, D. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, *86*(2), 75–85. https://doi.org/10.1007/BF02300500

Jørgensen, K. (2003). Problem solving: the essence of player action in computer games. In *Level Up Conference Proceedings*. Utrecht, Netherlands: DiGRA.

Just, M., & Carpenter, P. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480.

Juul, J. (2007). A certain level of abstraction (pp. 510–515). Presented at the Situated play, Japan. Retrieved from http://www.jesperjuul.net/text/acertainlevel/

Kagee, A., Allie, S., & Lesch, A. (2010). Effect of a course in research methods on scientific thinking among psychology students. *South African Journal of Psychology*, *40*(3), 272–281. https://doi.org/10.1177/008124631004000306

Kahnemann, D., & Tversky, A. (1979). Prospect theory: An analysis of decision making under risk. *Econometrica*, *47*, 263–291.

Karle, J., Watter, S., & Shedden, J. (2010). Task switching in video game players: Benefits of selective attention but not resistance to proactive interference. *Acta Psychologica*, *134*(1).

Khang, H., Kim, J., & Kim, Y. (2013). Self-traits and motivations as antecedents of digital media flow and addiction: The Internet, mobile phones, and video games. *Computers in Human Behavior, 29*, 2416–2424.

King, D., Delfabbro, P., & Griffiths, M. (2013). Chapter 82 – Video game addiction. In *Principles of Addiction. Comprehensive Addictive Behaviors and Disorders* (Vol. 1, pp. 819–825). Massachusetts, United States of America: Academic Press.

Kisiel, J., Rowe, S., Vartabedian, M., & Kopczak, C. (2012). Evidence for family engagement in scientific reasoning at interactive animal exhibits. *Science Education*, *96*(6), 1047–1070. https://doi.org/10.1002/sce.21036

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Sicence*, *12*(1), 1–48. Retrieved from http://onlinelibrary.wiley.com/doi/10.1207/s15516709cog1201_1/epdf

Klei Entertainment. (2013). *Don't Starve* [Video game].

Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, *64*(3), 141–152. https://doi.org/10.1024/1421-0185.64.3.141

Kuhn, D. (2009). Do students need to be taught how to reason? *Educational Research Review*, (4), 1–6. https://doi.org/10.1016/j.edurev.2008.11.001

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: what needs to develop to achieve skilled scientific thinking? *Cognitive Development*, *23*, 435–451. https://doi.org/10.1016/j.cogdev.2008.09.006

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction*, *9*(4), 285–327. https://doi.org/10.1016/S0959-4752(96)80001-6

Laskey, K. (2008). MEBN: A language for first-order Bayesian knowledge bases. *Artificial Intelligence*, *172*, 140–178. https://doi.org/10.1016/j.artint.2007.09.006

Lassiter, D., & Goodman, N. (2015). How many kinds of reasoning? Inference, probability, and natural language semantics. *Cognition*, *136*, 123–134.

Law, B., Atkins, S., Kirkpatrick, A., & Lomax, A. (2004). Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In *Proceedings of the 2004 symposium on Eye tracking research & applications* (pp. 41–48). New York, USA: ACM.

Lawson, A. (2000). The generality of hypothetico-deductive reasoning: Making scientific thinking explicit. *The American Biology Teacher*, *62*(7), 482–495. https://doi.org/10.1662/0002-7685(2000)062[0482:TGOHDR]2.0.CO;2

Lawson, A., & Daniel, E. (2011). Inferences of clinical diagnostic reasoning and diagnostic error. *Journal of Biomedical Informatics*, *44*, 402–412. https://doi.org/10.1016/j.jbi.2010.01.003

Lazonder, A., Hagemans, M., & de Jong, T. (2010). Offering and discovering domain information in simulation-based inquiry learning. *Learning and Instruction*, *20*, 511–520. https://doi.org/10.1016/j.learninstruc.2009.08.001

League system. (2016). [Wiki]. Retrieved 15 May 2016, from http://leagueoflegends.wikia.com/wiki/League_system

Lieberoth, A., & Roepstorff, A. (2015). Mixed methods in game research. Playing on strengths and countering weaknesses. In P. Lankoski & S. Björk (Eds.), *Game Research Methods* (pp. 271–290). ETC Press.

Lilienfeld, S., Ammirati, R., & David, M. (2012). Distinguishing science from pseudoscience in school psychology: science and scientific thinking as safeguards against human error. *Journal of School Psychology*, *50*(1), 7–36. https://doi.org/10.1016/j.jsp.2011.09.006

Linn, M., & Rice, M. (1979). A measure of scientific reasoning: the springs task. *Journal of Educational Measurement*, *16*(1), 55–58.

Lipsit, L., & Reese, H. (1981). *Desarrollo infantil*. Mexico: Trillas.

Loewenstein, G. (1994). The psychology of curiosity: a review and reinterpretation. *Psychological Bulletin*, *116*(1), 75–98.

Magno, C. (2011). Assessing the relationship of scientific thinking, self-regulation in research, and creativity in a measurement model. *The International Journal of Research and Review*, *6*(1), 17–47.

Maiberg, E. (2014, February 8). Blizzard eSports initiative will support your college gaming club. *Gamespot*. Retrieved from http://www.gamespot.com/articles/blizzard-esports-initiative-will-support-your-college-gaming-club/1100-6417615/

Martin, C., Chu, S., Johnson, D., Ochsner, A., Williams, C., & Steinkuehler, C. (2011). Ding! World of Warcraft well played, well researched. In *Well played 3.0* (pp. 226–245). Pittsburgh, USA: ETC Press.

Martin, J., & Olin, D. (1982). Improved methods for strategic analysis. *Policy Sciences*, *15*, 51–69.

Matchmaking. (2016). [Wiki]. Retrieved 15 May 2016, from http://dota2.gamepedia.com/Matchmaking

Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Austria. Retrieved from http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173

McCabe, D., & Castel, A. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, *107*, 343–352. https://doi.org/10.1016/j.cognition.2007.07.017

McClarty, K., Orr, A., Frey, P., Dolan, R., Vassileva, V., & McVay, A. (2012). *A literature review of gaming in education* (Gaming in Education). Pearson.

Michael, D., & Chen, S. (2006). *Serious games: Games that educate, train, and inform*. Canada: Thomson Course Technology PTR.

Miller, D. (1989). Matching strategies and strategy making: Process, content, and performance. *Human Relations*, *42*(3), 241–260.

Miller, G. (1983). Is scientific thinking different? *Bulletin of the American Academy of Arts and Sciences*, *36*(5), 26–37.

Miller, P. (2010, December 29). 2011: The year of eSports. *PCWorld*. Retrieved from http://www.pcworld.com/article/214432/2011_the_year_of_esports.html

Montealegre, R. (2007). La solución de problemas cognitivos. Una reflexión cognitiva sociocultural [Cognitive problem solving: A cognitive socio-cultural reflection]. *Avances En Psicología Latinoamericana*, *25*(2), 20–39. Retrieved from http://www.redalyc.org/articulo.oa?id=79925203

Mormann, T. (2012). A place for pragmatism in the dynamics of reason? *Studies in History and Philosophy of Science*, *43*, 27–37. https://doi.org/10.1016/j.shpsa.2011.10.004

Morse, A. F., Belpaeme, T., Cangelosi, A., & Floccia, C. (2011). Modeling U Shaped Performance Curves in Ongoing Development. In *CogSci 2011 Proceedings*. Boston, USA: Cognitive Science Society. Retrieved from http://www.tech.plym.ac.uk/SoCCE/staff/TonyBelpaeme/papers/Morse_CogSci2011.pdf

Nacke, L. E., Nacke, A., & Lindley, C. A. (2009). Brain training for silver gamers: Effects of age and game form on effectiveness, efficiency, self-assessment, and gameplay experience. *CyberPsychology & Behavior*, *12*(5), 493–499. https://doi.org/10.1089/cpb.2009.0013

Newell, A. (1966). On the analysis of human problem solving protocols. Presented at the International Symposium ofn Mathematical and Computational Methods in the Social Sciences, Carnegie Institute of Technology.

Newell, A., & Simon, H. (1972). *Human problem solving*. USA: Prentice-Hall.

Nichol, C., Szymczyk, A., & Hutchinson, J. (2014). Data first: Building scientific reasoning in AP chemistry via the concept development study approach. *Journal of Chemical Education, 91*, 1318–1325. https://doi.org/10.1021/ed500027g

Oei, A., & Patterson, M. (2013). Enhancing cognition with video games: A multiple game training study. *PLOS One*, *8*(3). https://doi.org/10.1371/journal.pone.0058546

Orvis, K., Horn, D., & Belanich, J. (2008). The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Computers in Human Behavior*, *24*(5), 2415–2433.

Osborne, T. (1998). *Aspects of Enlightenment. Social theories and the ethics of truth* (2005 ed.). London, United Kingdom: Taylor & Francis e-Library.

Papale, L. (2013). *Estetica dei videogiochi. Percorsi, evoluzioni, ibridazioni* [Video game aesthetics. Pathways, evolutions, and hybridisations]. Rome, Italy: Video Game Museum. Vigamus.

Patino, A., Romero, M., & Proulx, J.-N. (2016). Analysis of game and learning mechanics according to the learning theories. In *Proceedings of the 8th International Conference on Games and Virtual Worlds for Serious Applications*. Barcelona, Spain: IEEE. Retrieved from http://vsgames2016.com/proceedings/papers/pid1178402.pdf

Patterson, C. (1994). Delineation of separate brain regions used for scientific versus engineering modes of thinking. *Geochimica et Cosmochimica Acta*, *58*(15), 3321–3327.

Pereira, R., Wilwert, M., & Takase, E. (2016). Contributions of Sport Psychology to the Competitive Gaming: An Experience Report with a Professional Team of League of Legends. *International Journal of Applied Psychology 6*(2), 27-30.

Piaget, J. (1964). Development and learning. *Journal of Research in Science Teaching*, *2*, 176–186. https://doi.org/10.1002/tea.3660020306

Pike, K. (1967). Etic and Emic standpoints for the description of behavior. In *Language in Relation to a Unified Theory of the Structure of Human Behavior* (2nd ed., pp. 37–72). The Hague, Netherlands: Mouton & Co.

Piraksa, C., Srisawasdi, N., & Koul, R. (2014). Effect of gender on students' scientific reasoning ability: A case study in Thailand. *Procedia - Social and Behavioral Sciences*, *116*, 486–491. https://doi.org/10.1016/j.sbspro.2014.01.245

Pöllänen, T. (2014). *The influence of cognitive skills and team cohesion on player performance in Multiplayer Online Battle Arena*. University of Ljubljana, Slovenia. Retrieved from http://web-psychometrics.com/images/LoL.pdf

Poole, A., & Ball, L. (2005). Eye tracking in human-computer interaction and usability research: current status and future prospects. In *Encyclopedia of human computer interaction* (pp. 211–219). Pensilvania: Idea Group.

Poole, A., Ball, L., & Phillips, P. (2004). In search of salience: A response time and eye movement analysis of bookmark recognition. In P. Fincher, Markopolous, D., D. Moore, & R. Ruddle (Eds.), *People and Computers XVIII-Design for Life: Proceedings of HCI 2004.* London: Springer-Verlag Ltd.

Popper, B. (2013, September 30). Field of streams: how Twitch made video games a spectator sport. *The Verge*. Retrieved from http://www.theverge.com/2013/9/30/4719766/twitch-raises-20-million-esports-market-booming

Popper, K. (1959). *The logic of scientific discovery* (2002 ed.). New York: Routledge Classics.

Popper, K. (1962). *Conjectures and refutations. The Growth of Scientific Knowledge* (2014 ed.). USA: Routledge.

Popper, K. (1966). A realist view of logic, physics, and history. In *Objective Knowledge: An Evolutionary Approach* (1972 ed.). Oxford University Press.

Prensky, M. (2002). *What kids learn that's POSITIVE from playing video games*. Essay. Retrieved from http://www.marcprensky.com/writing/Prensky%20-%20What%20Kids%20Learn%20Thats%20POSITIVE%20From%20Playing%20Video%20Games.pdf

Prensky, M. (2005). In educational games, complexity matters. *Educational Technology*, *45*(4).

Prensky, M. (2006). *Don't bother me mom-I'm learning*. USA: Paragon House.

Pupil-labs. (2014). Pupil-Labs [Company page]. Retrieved from https://pupil-labs.com/

Pyper, B. (2012). Changing scientific reasoning and conceptual understanding in college students. In *Proceedings of the 2011 Physics Education Research Conference*. Omaha, USA: AIP Publishing.

Rabbitt, P. (1997). *Methodology of frontal and executive function*. United Kingdom: Psychology Press.

Riot Games. (2009). *League of Legends* [Ms Windows, OS X]. Riot Games. Retrieved from http://euw.leagueoflegends.com/

Roberge, J. (1970). A study of children's abilities to reason with basic principles of deductive reasoning. *American Educational Research Journal*, *7*(1), 583–596.

Robert, G., & Ross, I. (2016). R (Version 3.2.1) [Windows]. Retrieved from https://www.r-project.org/about.html

Rogers, S. (2010). *Level Up: A guide to great videogame design*. USA: John Wiley & Sons, Ltd.

Rosas, R., Grau, V., Salinas, M., & Correa, M. (2000). Más allá de Mortal Kombat: Diseño y evaluación de video juegos educativos para lenguaje y matemáticas del nivel básico [Beyond Mortal Kombat: Design and evaluation of educational video games for basic languages and mathematics]. *Psykhe*, *9*(2), 125–141.

Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., … Salinas, M. (2003). Beyond Nintendo: design and assessment of educational video games for first and second grade students. *Computers & Education*, *40*(1), 71–94. https://doi.org/10.1016/S0360-1315(02)00099-4

Ruphy, S. (2011). From Hacking's plurality of styles of scientific reasoning to 'foliated' pluralism: A philosophically robust form of ontologico-methodological pluralism. *Philosophy of Science*, *78*(5), 1212–1222. https://doi.org/0031-8248/2011/7805-0041$10.00

Russ, R., Coffey, J., Hammer, D., & Hutchison, P. (2009). Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking. *Science Education*, *93*(5), 875–891. https://doi.org/DOI 10.1002/sce.20320

Sadler, T. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*, *14*(5), 513–536. https://doi.org/10.1002/tea.20009

Scardamalia, M., & Bereiter, C. (2006). Knowledge building. Theory, pedagogy, and technology. In *Cambridge Handbook of the Learning Sciences* (pp. 97–118). New York, USA: Cambridge Handbook of the Learning Sciences.

Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *The Journal of the Learning Sciences*, *1*(2), 201–238.

Schoenbach, V. (1999). Inferencia causal [Causal inference]. In *Comprendiendo los Fundamentos de la Epidemiología* (pp. 269–287). Retrieved from http://www.epidemiolog.net/es/endesarrollo/InferenciaCausal.pdf

Schrader, P., & McCreery, M. (2008). The acquisition of skill and expertise in massively multiplayer online games. *Educational Technology Research and Development, 56*, 557–574. https://doi.org/10.1007/s11423-007-9055-4

Schum, D. (2001). *The evidential foundations of probabilistic reasoning*. Virgina, USA: John Wiley & Sons, Inc.

Schunn, C., & Anderson, J. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Sicence*, *23*(3), 337–370. https://doi.org/10.1207/s15516709cog2303_3

Scientific Method. (2015). In *Stanford Encyclopedia of Philosophy*.

Scientific Method. (2016). *Oxford Dictionary*. Retrieved from http://www.oxforddictionaries.com/definition/english/scientific-method

Seger, C. A. (1994). Implicit learning. Psychological bulletin, 115(2), 163.

Siegler, R. (1999). Strategic development. *Trends in Cognitive Sciences*, *3*(11), 430–435.

Siegler, R., & Atlas, M. (1976). Acquisition of formal scientific reasoning by 10- and 13-Year-Olds: Detecting Interactive Patterns in Data. *Journal of Educational Psychology, 68*(3), 360–370.

Sinclair, J. (1978). Experimental design deck for clinical research: A new learning resource in aid of scientific thinking. *Pediatrics*, *62*(2), 222–227.

Singer, D., & Revenson, T. (1996). *A Piaget primer. How a child thinks.* (Revised edition). Middlesex, England.: The Penguin Group.

Smith, S., Stibric, M., & Smithson, D. (2013). Exploring the effectiveness of commercial and custom-built games for cognitive training. *Computers in Human Behavior*, *29*, 2388–2393.

Soares, C., Hoga, L., Peduzzi, M., Sangaleti, C., Yonekura, T., & Silva, D. (2014). Integrative review: Concepts and methods used in nursing. *Revista Da Escola de Enfermagem Da USP*, *48*(2), 329–339. https://doi.org/10.1590/S0080-623420140000200020

Stebbins, R. (2001). *Exploratory research in the social sciences*. United States of America: Sage Publications, Inc.

Steinkuehler, C. (2004). Learning in Massively Multiplayer Online Games. In *Proceedings of the Sixth International Conference of the Learning Sciences* (pp. 521–528). New Jersey, USA.: Erlbaum.

Steinkuehler, C. (2006). Massively Multiplayer Online Video Gaming as participation in a discourse. *Mind, Culture, and Activity*, *13*(1), 38–52.

Steinkuehler, C. (2008). Cognition and literacy in massively multiplayer online games. In *Handbook of research on new literacies* (pp. 611–634). New York, USA: Lawrence Erlbaum Associate.

Steinkuehler, C. (2010). Video games and digital literacies. *Journal of Adolescent and Adult Literacy*, *54*(1), 61–63. https://doi.org/10.1598/JAAL54.1.7

Steinkuehler, C., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education Technologies, 17,* 530–543. https://doi.org/10.1007/s10956-008-9120-8

Tassi, P. (2012, December 20). 2012: The year of eSports. *Forbes*. Retrieved from http://www.forbes.com/sites/insertcoin/2012/12/20/2012-the-year-of-esports/#261ba3b240fc

Tassi, P. (2015, January 8). Second US College now offering ‘League of Legends’ scholarship. *Forbes*. Retrieved from http://www.forbes.com/sites/insertcoin/2015/01/08/second-us-college-now-offering-league-of-legends-scholarship/#6a57b69d35f2

Thinking Rabbit. (1982). *Sokoban* [Video game].

Thoron, A., & Myers, B. (2012). Effects of inquiry–based agriscience instruction on student scientific reasoning. *Journal of Agricultural Education, 52*(4), 156–170.

Torraco, R. (2005). Writing integrative literature reviews: guidelines and examples. *Human Resource Development Review, 4*(3), 356–367. https://doi.org/10.1177/1534484305278283

Toubes, A., Santos, H., Llosa, S., & Llomagno, C. (2006). *Revisión del concepto de educación no formal* [Revision of the concept of ‘non-formal education’]. Buenos Aires, Argentina: Facultad de Filosofía y Letras UBA.

Tran, N., Schneider, S., Duran, L., Conley, A., Richland, L., Burchinal, M., … Martinez, M. (2012). The effects of mathematics instruction using spatial temporal cognition on teacher efficacy and instructional practices. *Computers in Human Behavior, 28,* 340–349. https://doi.org/10.1016/j.chb.2011.10.003

Tversky, A., & Kahnemann, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Valve Corporation. (2013). *Dota 2* [Ms Windows, OS X, Linux]. Valve Corporation. Retrieved from www.dota2.com

van den Haak, M., De Jong, M., & Schellens, P. (2010). Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology*, *22*(5), 339–351. https://doi.org/10.1080/0044929031000

van Gog, T., Paas, F., van Merriënboer, J., & Witte, P. (2005). Uncovering the problem-solving process: cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, *11*(4), 2037–244. https://doi.org/10.1037/1076-898X.11.4.237

van Lehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, *47*, 513–539.

von der Malsburg, T. (2015). Saccades (Package for R) (Version 0.1-1) [Windows]. Germany. Retrieved from https://github.com/tmalsburg/saccades

Vygotsky, L. (1979). *El desarrollo de los procesos psicológicos superiores* [Development of higher cognitive functions]. Barcelona: Crítica.

Wagner, M. (2006). On the scientific relevance of eSports. In *Proceedings of the 2006 International Conference on Internet Computing & Conference on Computer Games Development*. Las Vegas, USA.: Department for Interactive Media and Educational Technology. Retrieved from https://www.researchgate.net/profile/Michael_Wagner12/publication/220968200_On_the_Scientific_Relevance_of_eSports/links/00b4952589870231be000000.pdf

Walther, B. (2003). Playing and gaming. Reflections and classifications. *Game Studies*, *3*(1). Retrieved from http://www.gamestudies.org/0301/walther/

Wellman, M., Estelle, J., Singh, S., Vorobeychik, Y., Kiekintveld, C., & Soni, V. (2005). Strategic interactions in a supply chain game. *Computational Intelligence*, *21*(1), 1–26.

Williams, P. (2015). Drawing the line: An anthropological study of Star Trek Online players.

Yang, P., Harrison, B., & Roberts, D. (2014). Identifying patterns in combat that are predictive of success in MOBA Games. In *Proceedings of the Foundations of Digital Games 2014 Conference*. Florida, USA.: DiGRA. https://doi.org/10.1109/CIG.2012.6374171

Yeh, K.-H., & She, H.-C. (2010). On-line synchronous scientific argumentation learning: Nurturing students' argumentation ability and conceptual change in science context. *Computers & Education*, *55*(2), 586–602. https://doi.org/10.1016/j.compedu.2010.02.020

Young, M., Slota, S., Cutter, A., Jalette, G., Mullin, G., Lai, B., … Yukhymenko, M. (2012). Our princess is in another castle a review of trends in serious gaming for education. *Review of Educational Research*, *82*(1), 61–89. https://doi.org/DOI: 10.3102/0034654312436980

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, *20*, 99–149.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223.

List of Tables

List of Figures

Appendix

Appendix A

List of the analytical categories derived from the fields “Concept Characteristics”, “Teleologies”, and “Associated Concepts”

Concept Characteristics

Characteristics	Frequency
Abductive	2
Abstraction	3
Analogical Reasoning	6
Analytical	10
Argumentation	8
Auxiliary Explanations	3
Auxiliary Hypotheses	1
Bayesian	1
Belief Biased	1
Categories(Taxonomies)	3
Causal Reasoning	9
Classification	3
Combinatorial Reasoning	3
Communication	4
Complex	5
Concept Generation	3
Concept Integration	2
Confirmation	1
Conservation	1
Context Sensitive	1
Correlational Reasoning(Variable Correlation)	7
Creative	2
Critical Reasoning	9

(continued)

Concept Characteristics (continued)

Characteristics	Frequency
Curiosity	1
Data Interpretation	2
Decision Making	3
Deductive	9
Discovery	1
Distributed Reasoning	1
Domain General	1
Domain Specific	1
Drawing Conclusions	1
Effect Size	1
Empirical Evidence	5
Epistemology	1
Evidence Based	8
Evidence Evaluation	27
Evidence Generation	17
Executive Function	7
Experiment Design	32
Experiment Evaluation	20
Explanatory	7
Exploratory	4
External Definition	3
Extrapolation	1
Falsification	7
Generalisation	1
Heuristic	1
Historical	2
Hypotheses Contrast	7
Hypothesis Creation	39
Hypothesis Testing	22

(continued)

Concept Characteristics (continued)

Characteristics	Frequency
Hypothetico-Deductive	7
Inductive	12
Inferring	8
Inquiry	9
Interpretation	4
Intervention Based	1
Knowledge Generation	5
Law/Principle Generation	2
Logical Thinking	6
Mathematical	1
Measuring	5
Metacognition	6
Modelisation	6
Model Revision	1
ND (Not Defined)	20
Non-Biased (Rational/Objective)	10
Non-Utilitarian	1
Not Collection of Facts	2
Observing	12
Operationalise	1
Pattern Detection	7
Prediction	6
Probabilistic Reasoning	10
Problem Identification	3
Problem Solving	9
Processual	1
Proportional Reasoning	5
Reliability	2
Replication	1

(continued)

Concept Characteristics (continued)

Characteristics	Frequency
Representational Ability	3
Rule Creation	1
Science (Domain)	2
Scientific Method	6
Scientific Writing	1
Separate Scientific and Non Scientific Data	3
Sceptic	2
Social	1
Statistical Reasoning	2
Strategy	1
Structural	1
Synthesis	2
Systematic(Controlled)	7
Theory-Evidence Coordination	6
Theory Construction	11
Theory Revision	4
Thinking Scientifically	5
Understanding	3
Variable Control	18
Variable Identification	8
Variable Isolation	5

Teleologies

Teleology	Frequency
Access Scientific Knowledge	1
Argumentation	1
Better Represent the World	1
Conceptual Development	3
Conceptual Learning	5
Coordinate Theory-Evidence	2
Correct Errors	1
Create Scientific Knowledge	6
Decision Making	3
Describe Nature	1
Develop Critical Thinking	1
Develop High-order Thinking	3
Develop Scientific Artifacts	1
Develop Scientific Attitude	2
Develop Scientific Understanding	1
Discover Laws and Theories	4
Distinguish Science from non-Science	1
Evidence-Based Practices	3
Evidence Evaluation	4
Hypothesis Testing	3
Improve Game Design	1
Improve Healthcare Practices	5
Inquiry	1
Know the Truth	1
Learn Sciences	11
Learn a Specific Science	1
Learn to Think	1
ND (Not Defined)	27

(continued)

Teleologies (continued)

Teleology	Frequency
Non-Teleological	1
Perform Better in School	2
Predictions	2
Problem Solving	8
Reason Scientifically	4
Research	10
Science Understanding	9
Scientific Communities	2
Scientific Development	12
Scientific Discovery	11
Scientific Literacy	2
Scientific Method	2
Scientific Practices	5
Survive the Scientific Publication World	1
Systematic Tackle Problems	1
Theory Development	3
Think Like a Scientists	8
Understand and Solve Scientific Issues	3
Understand and Solve Social Issues	2
Understand Complex Phenomena	3
Understand Covariation and Correlation	1
Understand Human Behaviour	1
Understand the Scientific Method	2
Understand the World	3

Associated Concepts

Associated Concept	Frequency
Analogical	1
Analytical	3
Causal Reasoning	6
Collaboration	1
Conceptual	3
Creativity	3
Critical Thinking	8
Decision Making	1
Determinism	1
Dialogical	1
Epistemology	3
Experimentation	3
Explanatory	2
Exploratory	3
Falsification	1
Forecasting (Prediction)	1
Formal Reasoning	1
Hypothetico-Deductive Reasoning	3
Integrative	1
Measurement	1
Metaphoric Thinking	1
Modelisation	4
Multi-Causal Reasoning	1
Paradigm	1
Prior Beliefs	1
Probes	1
Problem Solving	38
Research	5

(continued)

Associated Concepts (continued)

Associated Concept	Frequency
Science Laws	3
Scientific Achievement	1
Scientific Argumentation	13
Scientific Attitude	3
Scientific Communication	1
Scientific Communities	3
Scientific Discourse	1
Scientific Discovery	30
Scientific Induction	1
Scientific Inference	4
Scientific Inquiry	37
Scientific Invention	1
Scientific Knowledge	3
Scientific Literacy	12
Scientific Method	34
Scientific Practice	3
Scientific Revolution	3
Scientific Skills	2
Scientific Theories	9
Strategic	2
Styles of Scientific Reasoning	4
Theory of Science	1
Understanding	1

Appendix B

Auditor´s report

on

Carlos Mauricio Castaño Díaz: Conceptual Review on Scientific Reasoning and Scientific Thinking, Ludwig-Maximilians-Universität München

Auditor: Prof. Dr. Birgit Dorner, Department of Social Work München, Katholische Stiftungsfachhochschule München

The auditor is the first supervisor of Carlos Mauricio Castaño Díaz dissertation project **"**How Multiplayer Online Battle Arenas Foster Scientific Reasoning and Argumentation". Therefore I am familiar with the research project but during the supervisory process I did not conduct research myself from a similar theoretical perspective as the researcher and doctoral student, nor was I involved in the process of data gathering and analysis so I could look at the research processes from a certain distance and could evaluate critically all the process steps during the research.

Audit Trail

The research quality, the visibility, comprehensibility and acceptability of the study were assessed following the audit procedure developed by Akkerman, Admiraal, Brekelmans, & Oost (2008) for auditing quality of qualitative research.

This audit procedure forces the researcher to make explicit the steps planning the research project and substantiate the decisions during the process of data gathering and analysis. The process ensures the accuracy of data gathering, the reliability and validity of data analysis, furthermore the auditor questions to what extent the steps undertaken in the study are appropriate, make sense and are acceptable in relation to the methodological standards of the domains the study was conducted in.

This process involves a dialogical interaction between the writers and an auditor starting with an audit contract between auditor and auditee.

To give the auditor an orientation over the planned research steps a 'ReadmeFirst document' was created. It contains the research question, the conceptual framework the planned and used methods and the expected results. The auditee prepared an audit trail in which the whole procedure of data gathering and analysis is documented, including all raw data material, categorized and processed data in various steps, research tools like the various codes, network analysis e.g. and the findings using a shared Dropbox between the auditee and the auditor. So the latter could follow every step more or less simultaneously and see what was done when, how and why.

Data Collection was clearly documented, missing data was accurately explained.

During the audit process the auditor could ask for the specification of unclear sections in the research documentation to guarantee the comprehensibility and acceptability of the study.

The findings are understandable and accurately documented.

The final document is not part of this audit process as it will be part of the dissertation and therefore part of another evaluation process in which the auditor is involved and it seems not appropriate that the auditor carries out two different evaluation processes at the same time with different goals and in different roles.

Result

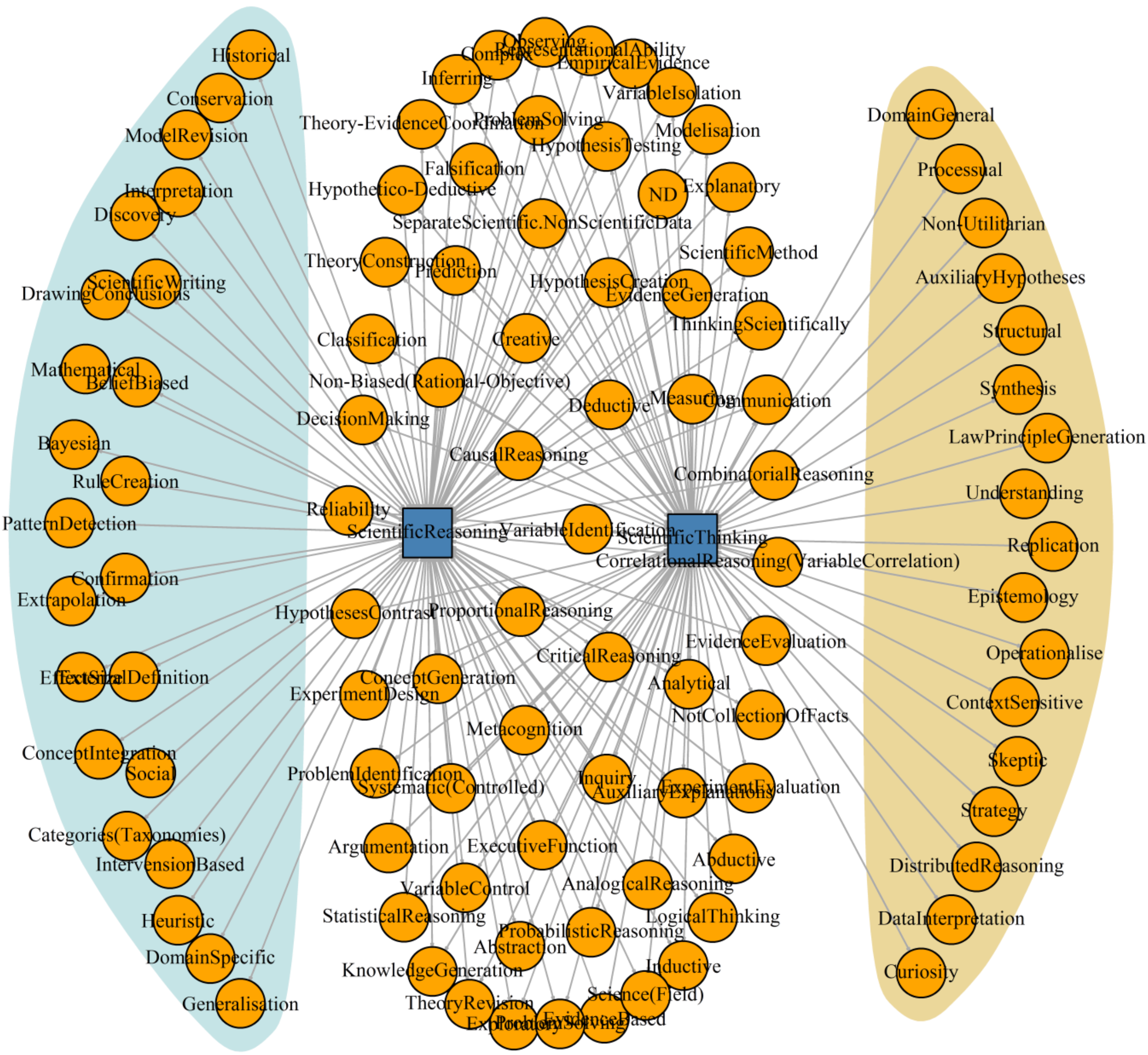
The decisions made by the researcher during this process were visible, well argued, and acceptable considering the rules of reliability and validity in research in the Learning Sciences and Game Studies. The auditor concludes that the overall quality of this research is satisfying in terms of the reliability and validity.

Munich, 16th August 2016

Appendix C

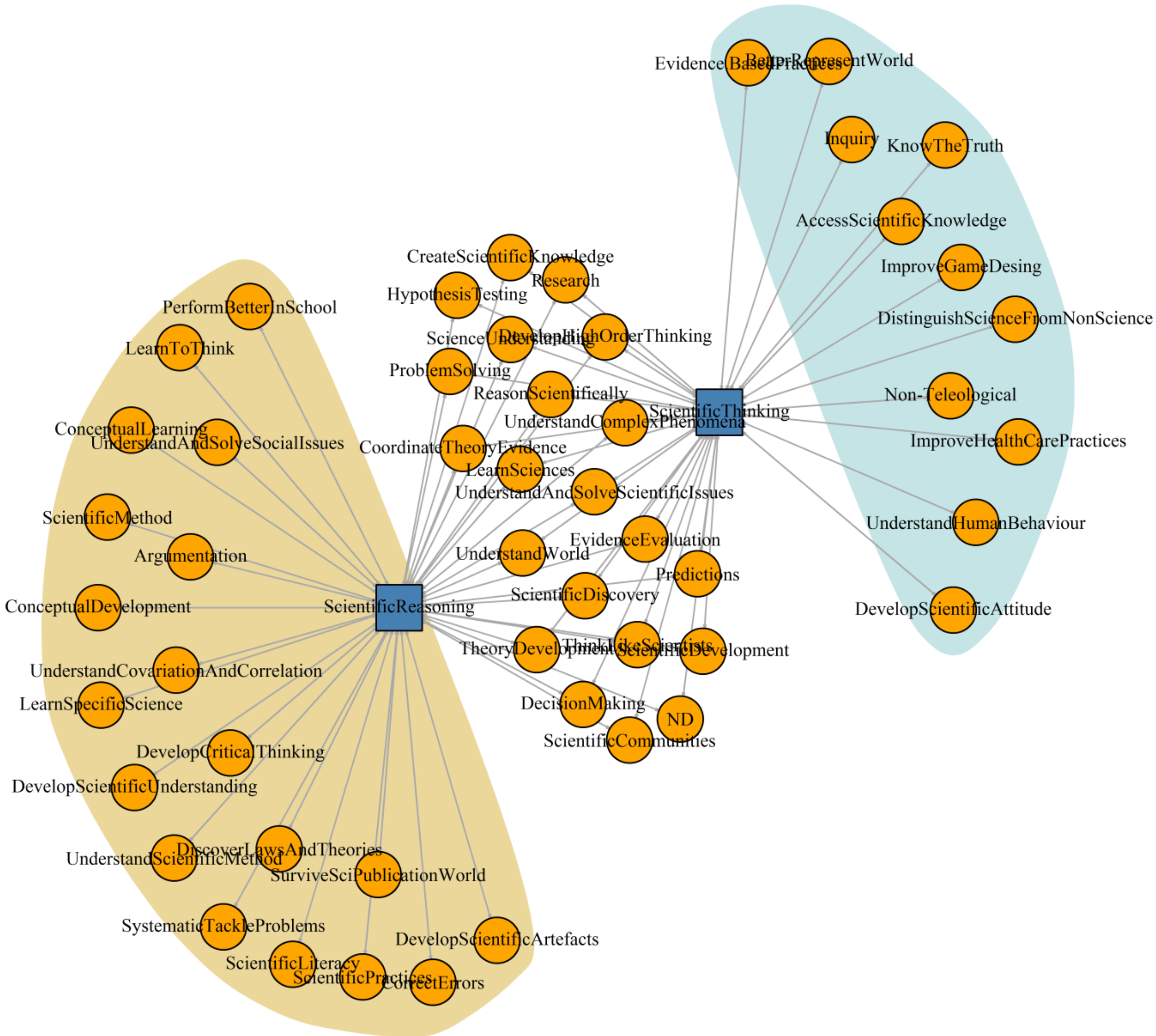
Graphs created using the 'igraph' package for R

Network analysis for the concepts of SR and ST based on their characteristics



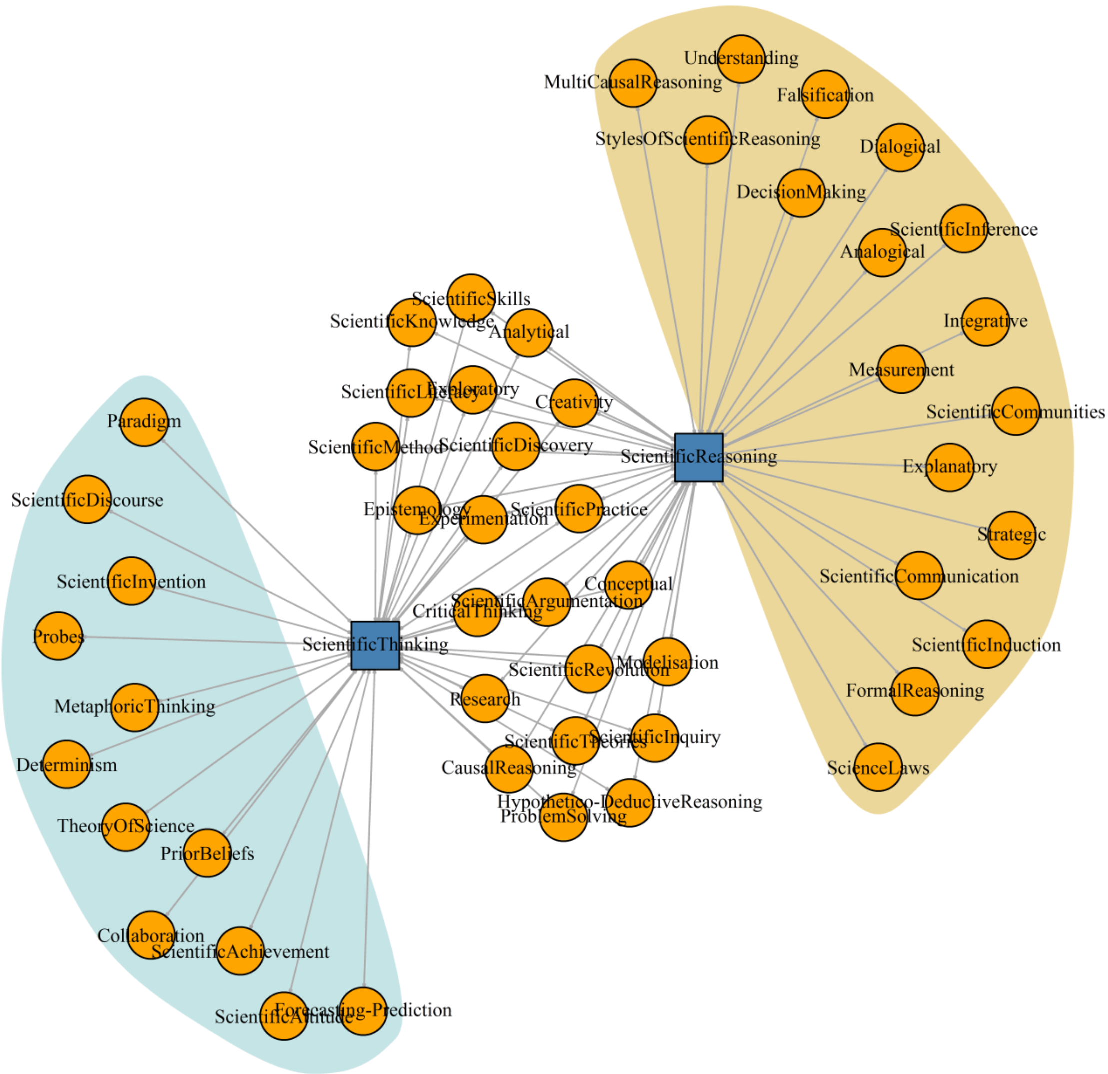
Depiction of the links between SR and ST using their Characteristics. The blue squares indicate the main concepts. The yellow circles indicate the Characteristics Associated. The light-blue area surrounding the yellow circles indicates the characteristics that are unique for SR. The light-yellow area surrounding the yellow circles indicates the characteristics that are unique for ST. The remaining characteristics are shared between the two concepts.

Network analysis for the concepts of SR and ST based on their Teleologies



Depiction of the links between SR and ST using their Teleologies. The blue squares indicate the main concepts. The yellow circles indicate the Teleologies associated. The light-yellow area surrounding the yellow circles indicates the Teleologies that are unique for SR. The light-blue area surrounding the yellow circles indicates the Teleologies that are unique for ST. The remaining Teleologies are shared between the two concepts.

Network analysis for the concepts of SR and ST based on their associated concepts



Depiction of the links between SR and ST using their Associated Concepts. The blue squares indicate the main concepts. The yellow circles indicate the Associated Concepts found in the research. The light-yellow area surrounding the yellow circles indicates the Associated Concepts that are unique for SR. The light-blue area surrounding the yellow circles indicates the Associated Concepts that are unique for ST. The remaining Associated Concepts are shared between SR and ST.

Appendix D

Extensive bibliography of the texts analysed in the research.

Abdulkarim, R., & Al Jadiri, A. (2012). The effect of cooperative learning group division based on multiple intelligences theory and previous achievement on scientific thinking skills development of ninth grade students in Oman. *European Journal of Social Sciences*, *27*(4), 553–569.

Abdullah, S., & Shariff, A. (2008). The effects of inquiry-based computer simulation with cooperative learning on scientific thinking and conceptual understanding of gas laws. *Eurasia Journal of Mathematics, Science and Technology Education*, *4*(4), 387–398.

Acar, O. (2014). Scientific reasoning, conceptual knowledge, & achievement differences between prospective science teachers having a consistent misconception and those having a scientific conception in an argumentation-based guided inquiry course. *Learning and Individual Differences*, *30*, 148–154. http://doi.org/10.1016/j.lindif.2013.12.002

Acar, O., & Patton, B. (2012). Argumentation and formal reasoning skills in an argumentation-based guided inquiry course. *Procedia-Social and Behavioral Sciences*, *46*, 4756–4760.

Adams, D., & Fuchs, M. (1985). The fusion of artistic and scientific thinking. *Art Education*, *38*(6), 22–24.

Adey, P., & Csapó, B. (2013). Chapter 1. developing and assessing scientific reasoning. In B. Csapó & G. Szabó (Eds.), *Framework for Diagnostic Assessment of Science* (pp. 17–53).

Allie, S., & Demaree, D. (2010). Toward meaning and scientific thinking in the traditional freshman laboratory: opening the 'Idea Space'. In *Proceedings of the Physics Education Research Conference*.

Al-Sharaf, A. (2013). Developing scientific thinking methods and applications in Islamic education. *Education*, *133*(3), 272–282.

Ames, E. (1934). Christianity and scientific thinking. *The Journal of Religion*, *14*(1), 4–12.

Amsel, E., Klaczynski, P., Johnson, A., Bench, S., Close, J., Sadler, E., & Walker, R. (2008). A dual-process account of the development of scientific reasoning: The nature and development of metacognitive intercession skills. *Cognitive Development*, *23*(4), 452–471. http://doi.org/10.1016/j.cogdev.2008.09.002

Anastas, J. (2014). When is research good evidence? Issues in reading research. *Clinical Social Work Journal*, *42,* 107–115. http://doi.org/10.1007/s10615-013-0452-3

Azarpira, N., Amini, M., Kojuri, J., Pasalar, P., Soleimani, M., Kahni, S., … Lankarani, K. (2012). Assessment of scientific thinking in basic science in the Iranian second national Olympiad. *BCM Research Notes*, *5*(1), 1–7.

Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development*, *2*(3), 201–221. http://doi.org/10.1111/j.1467-9507.1993.tb00014.x

Bartlett, S. F. (1959). Some problems of scientific thinking. *Ergonomics*, *2*(3), 229–238. http://doi.org/10.1080/00140135908930430

Beatty, E., & Thompson, V. (2012). Effects of perspective and belief on analytic reasoning in a scientific reasoning task. *Thinking and Reasoning*, *18*(4), 441–460. http://doi.org/10.1080/13546783.2012.687892

Blair, G. (1940). The validity of the Noll Test of scientific thinking. *Journal of Educational Psychology*, *3*(1), 53–59.

Blair, G., & Goodson, M. (1939). Development of scientific thinking through general science. *The School Review*, *47*(9).

Boland, L. (1994). Chapter 8. Scientific Thinking without Scientific Method: Two Views of Popper. In Backhouse (Ed.), *New Directions in Economic Methodology* (pp. 154–172). London, United Kingdom: Routledge.

Brigandt, I. (2010). Scientific reasoning is material inference: combining confirmation, discovery, and explanation. *International Studies in Philosophy of Science*, *24*(1), 31–43. http://doi.org/10.1080/02698590903467101

Bronowski, J. (2014). The nature of scientific reasoning. In *The iDeal Reader*. McGraw-Hill. Retrieved from http://www.public.iastate.edu/~bccorey/105%20Folder/The%20Nature%20of%20Sci.pdf

Brown, N., Furtak, E., Timms, M., Nagashima, S., & Wilson, M. (2010). The evidence-based reasoning framework: Assessing scientific reasoning. *Educational Assessment*, *15*(3), 123–141. http://doi.org/10.1080/10627197.2010.530551

Buck, N., & Ojemann, R. (1942). The relation between ability in scientific thinking and behavior in situations involving choice. *The Journal of Experimental Education*, *11*(2), 215–219.

Catton, P. (1999). Problems with the deductivist image of scientific reasoning. *Philosophy of Science*, *66*, S452–S473.

Cocking, R., Mestre, J., & Brown, A. (2000). New developments in the science of learning: using research to help students learn science and mathematics. *Journal of Applied Developmental Psychology*, *21*(1), 1–11.

Coletta, V., Jeffrey, P., & Steinert, J. (2011). fci normalized gain, scientific reasoning ability, thinking in physics, and gender effects. In *Proceedings of the 2011 Physics Education Research Conference* (Vol. 1413, pp. 23–26). Nebraska, USA. http://doi.org/10.1063/1.3679984

Collings, J. (1985). Scientific thinking through the development of formal operations: training in the cognitive restructuring aspect of field. *Research in Science and Technological Education*, *3*(2). http://doi.org/10.1080/0263514850030107a

Croker, S., & Buchanan, H. (2011). scientific reasoning in a real-world context: the effect of prior belief and outcome on children's hypothesis-testing strategies. *British Journal of Developmental Psychology*, *29*(3), 409–424. http://doi.org/10.1348/026151010X496906

Crowley, K., Callanan, M., Jipson, J., Galco, J., Topping, K., & Shrager, J. (2001). Shared scientific thinking in everyday parent-child activity. *Science Education*, *85*(6), 712–732. http://doi.org/10.1002/sce.1035

Dejonckheere, P., van de Keere, K., & Mestdagh, N. (2010). Training the scientific thinking circle in pre- and primary school. *The Journal of Education Research*, *103*, 1–16. http://doi.org/10.1080/00220670903228595

Dell'Acqua, P., & Pereira, L. (2004). common-sense reasoning as proto-scientific agent activity. *Journal of Applied Logic*, *2*(4), 385–407. http://doi.org/10.1016/j.jal.2004.07.002

Delrieux, C. (2001). The rôle of defeasible reasoning in the modelling of scientific research programmes. In *Proceedings of the IC-AI 2001 Conference* (pp. 861–868).

Derry, S., Levin, J., Osana, H., Jones, M., & Peterson, M. (2000). Fostering students' statistical and scientific thinking: lessons learned from an innovative college course. *American Educational Research Journal*, *37*(3), 747–773.

Dewey, J. (1910). Chapter 11. Empirical and scientific thinking. In *How we Think* (pp. 145–156).

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153–170. http://doi.org/10.1016/j.learninstruc.2012.05.003

Downing, E. (1928). The elements and safeguards of scientific thinking. *The Scientific Monthly*, *26*(3), 231–243.

Druyan, S., & Levin, I. (1996). The differential contribution of field-dependent and field-independent cognitive styles to sociocognitive transaction in promoting scientific reasoning. *International Journal of Behavioral Development*, *19*(4), 831–850.

Dunbar, K. (2000). How scientists think in the real world: implications for science education. *Journal of Applied Developmental Psychology*, *21*(1), 49–58.

Dunbar, K. (2001). Chapter 5. What scientific thinking reveals about the nature of cognition. In K. Crowley, C. Schunn, & T. Okada (Eds.), *Designing for Science: Implications from Everyday, Classroom, and Professional Settings* (pp. 115–140). Psychology Press.

Dunbar, K. (2009). The biology of physics: what the brain reveals about our understanding of the physical world. In *Proceedings of the 2009 Physics Education Research Conference* (pp. 15–18). Ann Arbor, USA: American Institute of Pliysics.

Dunbar, K., & Klahr, D. (2012). Chapter 35. Scientific thinking and reasoning. In K. Holyoak & R. Morrison (Eds.), *Oxford Handbook of Thinking and Reasoning* (2012 ed.). New York, USA: Oxford University Press.

Echevarria, M. (2003). anomalies as a catalyst for middle school students' knowledge construction and scientific reasoning during science inquiry. *Journal of Educational Psychology*, *95*(2), 357–374.

Eflin, J., & Kite, M. (1996). Teaching scientific reasoning through attribution theory. *Teaching of Psychology, 23*(2), 87–91. http://doi.org/10.1207/s15328023top2302_3

El-Nasr, M., & Maygoli, T. (2011). Understanding the underlying patterns: Teaching scientific thinking through building games. *International Journal of Science in Society*, *2*(1), 17–27.

Engelhart, M., & Lewis, H. (1941). An attempt to measure scientific thinking. *Educational and Psychological Measurement*, *1*(1), 289–293.

Etkina, E., Karelina, A., & Villasenor, M. (2007). studying transfer of scientific reasoning abilities. In *2006 Physics Education Research Conference* (pp. 81–84). New York, USA: AIP Publishing. http://doi.org/http://dx.doi.org/10.1063/1.2508696

Faulkner, D., Joiner, R., Littleton, K., Miell, D., & Thompson, L. (2000). The mediating effect of task presentation on collaboration and children's acquisition of scientific reasoning. *European Journal of Psychology of Education*, *15*(4), 417–430. http://doi.org/10.1007/BF03172985

Fender, J., & Crowley, K. (2007). How parent explanation changes what children learn from everyday scientific thinking. *Journal of Applied Developmental Psychology*, *28*(3), 189–210. http://doi.org/10.1016/j.appdev.2007.02.007

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., … Eberle, J. (2014). Scientific reasoning and argumentation: advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, *2*(3), 28–45. http://doi.org/http://dx.doi.org/10.14786/flr.v2i2.96

Fraser, J., Shane-Simpson, C., & Asbell-Clarke, J. (2014). youth science identity, science learning, and gaming experiences. *Computers in Human Behavior*, *41*, 523–532. http://doi.org/10.1016/j.chb.2014.09.048

Freeman, J. (2003). Scientific reasoning in gifted children. In P. Csermely & L. Lederman (Eds.), *Science Education: Talent Recruitment and Public Understanding* (pp. 17–30). Ios Pr Inc.

Furuya, S., Shin, M., Kamei, E., & Sentoku, E. (2012). Formation of active attitude for learning and of habits of scientific thinking by project based team learning at Kanazawa Institute of Technology. In *Proceedings of the 8th International CDIO Conference*. Brisbane, Australia.

Gibbs, A., & Lawson, A. (1992). The nature of scientific thinking as reflected by the work of biologists and by biology textbooks. *The American Biology Teacher*, *54*(3), 137–152.

Giere, R. (2001). A new framework for teaching scientific reasoning. *Argumentation*, *15*, 21–33. http://doi.org/10.1023/A:1007880010975

Gleason, M., & Schauble, L. (1999). parents' assistance of their children's scientific reasoning. *Cognition and Instruction*, *17*(4).

Gower, B. (1997). Henri Poincare and Bruno de Finetti: conventions and scientific reasoning. *Studies in History and Philosophy of Sciences*, *28*(4), 657–679.

Greenhoot, A., Semb, G., Colombo, J., & Schreiber, T. (2004). Prior beliefs and methodological concepts in scientific reasoning. *Applied Cognitive Psychology*, *18*, 203–221. http://doi.org/10.1002/acp.959

Hacking, I. (1992). Statistical language, statistical truth and statistical reason: the self-authentification of a style of scientific reasoning. *The Social Dimensions of Science*, *3*.

Hacking, I. (2012). Language, truth and reason' 30 years later. *Studies in History and Philosophy of Science*, *43*, 599–609. http://doi.org/10.1016/j.shpsa.2012.07.002

Harre, R. (2004). Chapter 3. Properties and images. In D. Rothbart (Ed.), *Modeling: Gateway to the Unknown* (Vol. 1, pp. 29–45). Elsevier Inc.

Herring, J. (1918). Measurement of some abilities of scientific thinking. *The Journal of Educational Psychology*, *9*(10), 535–558.

Herring, J. (1919). Derivation of a scale to measure abilities in scientific thinking. *The Journal of Educational Psychology*, *10*(9), 417–132.

Hogan, K., & Fischerkeller, J. (2000). Chapter 5. Dialogue as data: assessing students' scientific reasoning with interactive protocols. In J. Mintzes, J. Wandersee, & J. Novak (Eds.), *Assessing Science Understanding: A Human Constructivist View* (pp. 96–129). Academic Press.

Hogan, K., Natasie, B., & Presley, M. (1999). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cognition and Instruction*, *17*(4), 379–432.

Holmes, J. (2013). Chapter 15. Using controversies to teach scientific thinking in psychology: topics and issues. In D. Dunn, R. Gurung, K. Naufel, & J. Wilson (Eds.), *Controversy in the Psychology Classroom: Using Hot Topics to Foster Critical Thinking* (pp. 245–256). American Psychological Association.

Holmes, J., & Beins, B. (2009). Psychology is a science: at least some students think so. *Teaching of Psychology*, *36*, 5–11. http://doi.org/10.1080/00986280802529350

Holmes, N., & Bonn, D. (2013). Doing science or doing a lab? Engaging students with scientific reasoning during physics lab experiments. In *Proceedings of the 2013 Physics Education Research Conference*. Portland USA. http://doi.org/10.1119/perc.2013.pr.034

Hopkins, J. (1996). Psychoanalytic and scientific reasoning. *British Journal of Psychotherapy*, *13*(1), 86–105.

Howes, M. (2012). Managing salience: The importance of intellectual virtue in analyses of biased scientific reasoning. *Hypatia*, *27*(4), 736–754.

Hsieh, W.-M., Cui, L., & Sharma, P. (2008). Facilitating critical and scientific thinking using group-based problem solving activities in a large undergraduate course. In *Selected Research, Development, and Papers on the Practice of Educational Communications and Technology* (pp. 124–133). Orlando, USA.: The Research and Theory Division.

Huber, F. (2005). Subjective probabilities as basis for scientific reasoning. *British Journal of Philosophy of Science*, (56), 101–116. http://doi.org/10.1093/phisci/axi105

Hyrkäs, K., Åstedt-Kurki, P., & Paunonen, M. (1999). Finnish nurse teachers' perceptions of their 'scientific thinking' and its development during master's-level teacher education at university. *Scandinavian Journal of Caring Science*, *13*, 129–133.

Kagee, A., Allie, S., & Lesch, A. (2010). Effect of a course in research methods on scientific thinking among psychology students. *South African Journal of Psychology*, *40*(3), 272–281. http://doi.org/10.1177/008124631004000306

Kim, T. (2013). scientific reasoning and methods in urban planning. *International Regional Science Review*, *36*(1), 36–43. http://doi.org/0.1177/0160017612441365

Kind, P. (2013). Establishing assessment scales using a novel disciplinary rationale for scientific reasoning. *Journal of Research in Science Teaching*, *50*(5), 530–560. http://doi.org/10.1002/tea.21086

Kisiel, J., Rowe, S., Vartabedian, M., & Kopczak, C. (2012). Evidence for family engagement in scientific reasoning at interactive animal exhibits. *Science Education*, *96*(6), 1047–1070. http://doi.org/10.1002/sce.21036

Klaczynski, P. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: a two-process approach to adolescent cognition. *Child Development*, *71*(5), 1347–1366. http://doi.org/0009-3920/2000/7105-0019

Klaczynski, P., & Narasimham, G. (1998). Development of scientific reasoning biases: cognitive versus ego-protective explanations. *Developmental Psychology*, *34*(1), 175–187.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Sicence*, *12*(1), 1–48. Retrieved from http://onlinelibrary.wiley.com/doi/10.1207/s15516709cog1201_1/epdf

Koerber, S., & Sodian, B. (2009). Reasoning from graphs in young children: preschoolers' ability to interpret covariation data from graphs. *Journal of Psychology of Science and Technology, 2*(2), 73–85. http://doi.org/10.1891/1939-7054.2.2.73

Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, *64*(3), 141–152. http://doi.org/10.1024/1421-0185.64.3.141

Konstantinodou, A., Cervero, J., & Castells, M. (2009). Argumentation and scientific reasoning: the 'double hierarchy' argument. In G. Çakmakci & M. Taşar (Eds.), *Contemporary Science Education Research:Scientific Literacy and Social Aspects of Science*.

Kuhn, D. (1995). Microgenetic study of change: What has it told us? *Psychological Science*, *6*(3), 133–139.

Kuhn, D. (2009). Do students need to be taught how to reason? *Educational Research Review*, (4), 1–6. http://doi.org/10.1016/j.edurev.2008.11.001

Kuhn, D., & Dean, D. (2004). Connecting scientific reasoning and causal inference. *Journal of Cognition and Development*, *5*(2), 261–288. http://doi.org/10.1207/s15327647jcd0502_5

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development*, *23*, 435–451. http://doi.org/10.1016/j.cogdev.2008.09.006

Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development*, *1*, 113–129.

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction*, *9*(4), 285–327. http://doi.org/10.1016/S0959-4752(96)80001-6

Laius, A., & Rannikmäe, M. (2011). Impact on student change in scientific creativity and socio-scientific reasoning skills from teacher collaboration and gains from professional in-service. *Journal of Baltic Science Education*, *10*(2), 127–137.

Lawson, A. (2000). The generality of hypothetico-deductive reasoning: making scientific thinking explicit. *The American Biology Teacher*, *62*(7), 482–495. http://doi.org/10.1662/0002-7685(2000)062[0482:TGOHDR]2.0.CO;2

Lazonder, A., Hagemans, M., & de Jong, T. (2010). Offering and discovering domain information in simulation-based inquiry learning. *Learning and Instruction*, *20*, 511–520. http://doi.org/10.1016/j.learninstruc.2009.08.001

Lazonder, A., & Kamp, E. (2012). Bit by bit or all at once? Splitting up the inquiry task to promote children's scientific reasoning. *Learning and Instruction*, *22*(6), 458–464. http://doi.org/10.1016/j.learninstruc.2012.05.005

Legare, C. (2014). The contributions of explanation and exploration to children's scientific reasoning. *Child Development Perspectives*, *8*(2), 101–106.

León, P., & Castells, M. (2001). Switching from everyday facts to scientific thinking. In *Developing Formal Developing Formal Thinking in Physics*. Retrieved from http://www. fisica. uniud. it/girepseminar2001/CS06/LEON_01_FIN AL. pdf

Lilienfeld, S. (2010). Can psychology become a science? *Personal and Individual Differences*, *49*, 281–288.

Lilienfeld, S., Ammirati, R., & David, M. (2012). Distinguishing science from pseudoscience in school psychology: Science and scientific thinking as safeguards against human error. *Journal of School Psychology*, *50*(1), 7–36. http://doi.org/10.1016/j.jsp.2011.09.006

Linn, M., & Rice, M. (1979). A measure of scientific reasoning: The springs task. *Journal of Educational Measurement*, *16*(1), 55–58.

Luce, M., Callanan, M., & Smilovic, S. (2013). Links between parents' epistemological stance and children's evidence talk. *Developmental Psychology*, *49*(3), 454–461. http://doi.org/10.1037/a0031249

Magnani, L. (2002). Conjectures and manipulations: External representations in scientific reasoning. *Mind & Society*, *3*(1), 9–31. http://doi.org/10.1007/BF02511863

Magno, C. (2011). Assessing the relationship of scientific thinking, self-regulation in research, and creativity in a measurement model. *The International Journal of Research and Review*, *6*(1), 17–47.

Magnusson, S., & Palincsar, A. (2005). Chapter 10. Teaching to promote the development of scientific knowledge and reasoning about light at the elementary school level. In J. Bransford & S. Donovan (Eds.), *How students learn: History, Mathematics, and Science in the Classroom* (pp. 421–474). The National Academic Press.

Mauldin, R. (2012). A novel approach to teaching scientific reasoning to future journalists: An intellectual framework for evaluating press reports about scientific research. *Science Communication*, *34*(2), 283–291. http://doi.org/10.1177/1075547011427976

Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: assessment and relations with cognitive abilities. *Learning and Instruction*, *29*, 43–55. http://doi.org/http://dx.doi.org/10.1016/j.learninstruc.2013.07.005

McCabe, D., & Castel, A. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, *107*, 343–352. http://doi.org/10.1016/j.cognition.2007.07.017

McCarty, W. (2007). Being reborn: The humanities, computing and styles of scientific reasoning. In *New Technology in Medieval and Renaissance Studies* (pp. 1–27).

Miller, G. (1983). Is scientific thinking different? *Bulletin of the American Academy of Arts and Sciences*, *36*(5), 26–37.

Mormann, T. (2012). A place for pragmatism in the dynamics of reason? *Studies in History and Philosophy of Science*, *43*, 27–37. http://doi.org/10.1016/j.shpsa.2011.10.004

Morris, B., Croker, S., Masnick, A., & Zimmerman, C. (2012). Chapter 4. The emergence of scientific reasoning. In *Current Topics in Children's Learning and Cognition* (pp. 61–82). INTECH. Open Science, Open Minds.

Morris, B., Croker, S., Zimmerman, C., Gill, D., & Romig, C. (2013). Gaming science: The 'gamification' of scientific thinking. *Frontiers in Psychology*, *4*, 1–16. http://doi.org/10.3389/fpsyg.2013.00607

Nichol, C., Szymczyk, A., & Hutchinson, J. (2014). Data first: building scientific reasoning in AP chemistry via the concept development study approach. *Journal of Chemical Education*, *91*, 1318–1325. http://doi.org/10.1021/ed500027g

Olshin, B. (2007). Scientific thinking and modernity meet traditional culture. In *Proceedings of the Second Advanced International Colloquium on Building the Scientific Mind*. Cancouver, Canada. Retrieved from http://www. learndev. org/dl/BtSM2007/BenjaminOlshin. pdf

Osbeck, L. (2014). Scientific reasoning as sense-making: Implications for qualitative inquiry. *Qualitative Psychology*, *1*(1), 34–36. http://doi.org/10.1037/qup0000004

Patterson, C. (1994). Delineation of Separate Brain Regions Used for Scientific Versus Engineering Modes of Thinking. *Geochimica et Cosmochimica Acta*, *58*(15), 3321–3327.

Penner, D., & Klahr, D. (1996). When to trust the data: Further investigations of system error in a scientific reasoning task. *Memory and Cognition*, *25*(5), 655–668.

Piekny, J., & Maehler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology*, *31*, 153–179. http://doi.org/10.1111/j.2044-835X.2012.02082.x

Pietarinen, A.-V., & Bellucci, F. (2014). New light on Peirce's conceptions of retroduction, deduction, and scientific reasoning. *International Studies in the Philosophy of Science*, *28*(4), 353–373. http://doi.org/10.1080/02698595.2014.979667

Piraksa, C., Srisawasdi, N., & Koul, R. (2014). Effect of gender on students' scientific reasoning ability: A case study in Thailand. *Procedia - Social and Behavioral Sciences*, *116*, 486–491. http://doi.org/10.1016/j.sbspro.2014.01.245

Pitkälä, K., Mäntyranta, T., Strandberg, T., Mäkelä, M., Vanhanen, H., & Varonen, H. (2000). Evidence-based medicine - how to teach critical scientific thinking to medical undergraduates. *Medical Teacher*, *2*(1).

Popper, K. (1959). *The Logic of Scientific Discovery* (2002 ed.). New York: Routledge Classics.

Popper, K. (1962). Conjectures and refutations. The growth of scientific knowledge (2014 ed.). USA: Routledge.

Popper, K. (1966). A realist view of logic, physics, and history. In *Objective Knowledge: An Evolutionary Approach* (1972 ed.). Oxford University Press.

Pratt, J. (1934). Buddhism and scientific thinking. *The Journal of Religion*, *14*(1), 13–24.

Psycharis, S. (2013). Examining the effect of the computational models on learning performance, scientific reasoning, epistemic beliefs and argumentation: An implication for the STEM agenda. *Computers & Education*, *68*, 253–265. http://doi.org/10.1016/j.compedu.2013.05.015

Pyper, B. (2012). Changing scientific reasoning and conceptual understanding in college students. In *Proceedings of the 2011 Physics Education Research Conference*. Omaha, USA: AIP Publishing.

Rieppel, O. (2013). Styles of scientific reasoning: Adolf Remane (1898–1976) and the German evolutionary synthesis. *Journal of Zoological Systematics and Evolutionary Research*, *51*(1), 1–12. http://doi.org/10.1111/jzs.12003

Ruffman, T., Perner, J., Olson, D., & Doherty, M. (1993). Reflecting on scientific thinking: children’s understanding of the hypothesis-evidence relation. *Child Development*, *64*(6), 1617–1636.

Ruphy, S. (2011). From Hacking’s plurality of styles of scientific reasoning to ‘foliated’ pluralism: A philosophically robust form of ontologico-methodological pluralism. *Philosophy of Science*, *78*(5), 1212–1222. http://doi.org/0031-8248/2011/7805-0041$10.00

Russ, R., Coffey, J., Hammer, D., & Hutchison, P. (2009). Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking. *Science Education*, *93*(5), 875–891. http://doi.org/DOI 10.1002/sce.20320

Sadaghiani, H. (2010). Scientific reasoning for pre-service elementary teachers. In *Proceedings of the 2010 Physics Education Research Conference*. Portland, USA: American Institute of Phiysics.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, *32*(1), 102–119.

Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *The Journal of the Learning Sciences*, *1*(2), 201–238.

Schlaretzki, W. (1960). Scientific reasoning and the Summum Bonum. *Philosophy of Science*, *27*(1), 48–57.

Schmaltz, R., & Lilienfeld, S. (2014). Hauntings, homeopathy, and the Hopkinsville Goblins: using pseudoscience to teach scientific thinking. *Frontiers in Psychology*, *5*, 1–5. http://doi.org/10.3389/fpsyg.2014.00336

Schunn, C., & Anderson, J. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science*, *23*(3), 337–370. http://doi.org/10.1207/s15516709cog2303_3

She, H.-C., & Lee, C.-Q. (2008). SCCR digital learning system for scientific conceptual change and scientific reasoning. *Computers & Education*, *51*(2), 724–742. http://doi.org/doi:10.1016/j.compedu.2007.07.009

She, H.-C., & Liao, Y.-W. (2010). Bridging scientific reasoning and conceptual change through adaptive web-based learning. *Journal of Research in Science Teaching*, *47*(1), 91–119. http://doi.org/10.1002/tea.20309

Sher, D. (1992). A computational normative theory of scientific evidence. *International Journal of Approximate Reasoning*, *6*, 505–524.

Siegler, R., & Atlas, M. (1976). Acquisition of formal scientific reasoning by 10- and 13-year-olds: Detecting interactive patterns in data. *Journal of Educational Psychology*, *68*(3), 360–370.

Simon, H. (1973). Does scientific discovery have a logic? *Philosophy of Science*, *40*(4), 471–480. Retrieved from http://www.jstor.org/stable/186282

Simonetta, A. (2003). Chapter 1. The beginnings of Greek scientific thinking. In *Short History of Biology from the Origins to the 20th Century* (pp. 5–25). Firenze University Press.

Sinclair, J. (1978). Experimental design deck for clinical research: A new learning resource in aid of scientific thinking. *Pediatrics*, *62*(2), 222–227.

Smith, J. (1988a). Inconsistency and scientific reasoning. *Studies in History and Philosophy of Science Part A*, *19*(4), 429–445.

Smith, J. (1988b). Scientific reasoning or damage control: Alternative proposals for reasoning with inconsistent representations of the world. In *Proceedings of the Biennial Meeting of the Philosophy of Science Association* (pp. 241–248). The University of Chicago Press.

Songer, N. (1989). Technological tools for scientific thinking and discovery. *Journal of Reading, Writing, and Learning Disabilities International*, *5*(1), 23–41. http://doi.org/10.1080/0748763890050103

Steinkuehler, C., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education Technologies*, *17*, 530–543. http://doi.org/10.1007/s10956-008-9120-8

Stevens, C., & Witkow, M. (2014). Training scientific thinking skills: Evidence from an MCAT2015-Aligned classroom module. *Teaching of Psychology*, *41*(2), 115–121. http://doi.org/10.1177/0098628314530341

Strom, A., & Barolo, S. (2011). Using the game of Mastermind to teach, practice, and discuss scientific reasoning skills. *PLoS Biology*, *9*(1), e1000578. http://doi.org/doi:10.1371/journal.pbio.1000578

Tajudin, N. 'ain, Saad, N. S., Rahman, N., Yahaya, A., Alimon, H., Dollah, M., & Mustaman, M. (2012). Mapping the level of scientific reasoning skills to instructional methodologies among Malaysian science-mathematics-engineering undergraduates. In *American Institute of Physics Conference Series* (pp. 262–267). Bandung, Indonesia: AIP. http://doi.org/10.1063/1.4724151

Thitima, Gamlunglert, & Sumalee, Chaijaroen. (2012). Scientific thinking of the learners learning with the knowledge construction model enhancing scientific thinking. *Procedia - Social and Behavioral Sciences*, *46*, 3771–3775. http://doi.org/10.1016/j.sbspro.2012.06.144

Thoron, A., & Myers, B. (2012). Effects of inquiry–based Agriscience instruction on student scientific reasoning. *Journal of Agricultural Education*, *52*(4), 156–170.

Tibbetts, G. G. (2013). 2 - Elements of scientific thinking: Skepticism, careful reasoning, and exhaustive evaluation are all vital. In *How the Great Scientists Reasoned* (pp. 5–24). Oxford: Elsevier. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780123984982000023

Timmerman, B., & Strickland, D. (2009). Faculty should consider peer review as a means of improving students' scientific reasoning skills. *Journal of the South Carolina Academy of Science*, *7*(1), 1–8.

Timmerman, B., Strickland, D., Johnson, R., & Payne, J. (2011). Development of a 'universal' rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education*, *36*(5), 509–547. http://doi.org/10.1080/02602930903540991

Tran, N., Schneider, S., Duran, L., Conley, A., Richland, L., Burchinal, M., … Martinez, M. (2012). The effects of mathematics instruction using spatial temporal cognition on teacher efficacy and instructional practices. *Computers in Human Behavior*, *28*, 340–349. http://doi.org/10.1016/j.chb.2011.10.003

Trickett, S., & Trafton, G. (2007). 'What if...': The use of conceptual simulations in scientific reasoning. *Cognitive Science*, *31*, 843–875.

Tytler, R., & Peterson, S. (2004). From ‘try it and see’ to strategic exploration: characterizing young children’s scientific reasoning. *Journal of Research in Science Teaching*, *41*(1), 94–118. http://doi.org/10.1002/tea.10126

van de Keere, K., Mestdagh, N., & Dejonckheere, P. (2007). Children learn by means of the scientific thinking circle. In N. Hall (Ed.), *The Proceedings of the 17th Annual Conference of the European Teacher Education Network* (pp. 56–60). University of Greenwich.

Voskoglou, M. (2013). An application of fuzzy sets to scientific thinking. In *Proceedings of the The 6th International Conference on Information Technology*. Las Vegas, USA.

Weld, J., Stier, M., & McNew-Birren, J. (2011). The development of a novel measure of scientific reasoning growth among college freshmen: The constructive inquiry science reasoning skills Test. *Journal of College Science Teaching*, *40*(4), 101–107.

Williams, R. (2010). Gesture in everyday scientific reasoning and explanation. In *Proceedings of the 4th Conference of the International Society for Gesture Studies* (pp. 25–30). Frankfurt, Germany.

Yeh, K.-H., & She, H.-C. (2010). On-line synchronous scientific argumentation learning: nurturing students’ argumentation ability and conceptual change in science context. *Computers & Education*, *55*(2), 586–602. http://doi.org/10.1016/j.compedu.2010.02.020

Zeineddin, A., & Abd-El-Khalick, F. (2008). On coordinating theory with evidence: The role of epistemic commitments in scientific reasoning among college students. *Eurasia Journal of Mathematics, Science & Technology Education*, *4*(2), 153–168.

Zeineddin, A., & Abd-El-Khalick, F. (2010). Scientific reasoning and epistemological commitments: Coordination of theory and evidence among college science students.

Journal of Research in Science Teaching, *47*(9), 1064–1093. http://doi.org/10.1002/tea.20368

Zhang, J., Chen, Q., & Reid, D. (2000). Simulation-based scientific discovery learning: A research on the effects of experimental support and learners' reasoning ability. In *Proceedings IFIP World Computer Congress, ICEUT 2000, Educational Uses of Communication and Information Technologies*. Beijing, China.

Zimbardi, K., Bugarcic, A., Colthorpe, K., Good, J., & Lluka, L. (2013). A set of vertically integrated inquiry-based practical curricula that develop scientific thinking skills for large cohorts of undergraduate students. *Advances in Physiology Education*, *37*, 303–315. http://doi.org/10.1152/advan.00082.2012.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, *20*, 99–149.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223.

Appendix E

Auditor´s report

on

Carlos Mauricio Castaño Díaz: Scientists in the Battle Ground: Scientific Reasoning and Strategy Making in Multiplayer Online Battle Arenas, Ludwig-Maximilians-Universität München

Auditor: Prof. Dr. Birgit Dorner, Department of Social Work München, Katholische Stiftungsfachhochschule München

The auditor is the first supervisor of Carlos Mauricio Castaño Díaz dissertation project **“**How Multiplayer Online Battle Arenas Foster Scientific Reasoning and Argumentation”. Therefore I am familiar with the research project but during the supervisory process I did not conduct research myself from a similar theoretical perspective as the researcher and doctoral student, nor was I involved in the process of data gathering and analysis so I could look at the research processes from a certain distance and could evaluate critically all the process steps during the research.

Audit Trail

The research quality, the visibility, comprehensibility and acceptability of the study were assessed following the audit procedure developed by Akkerman, Admiraal, Brekelmans, & Oost (2008) for auditing quality of explorative, qualitative and mixed-methods research.

This audit procedure forces the researcher to make explicit the steps planning the research project and substantiate the decisions during the process of data gathering and analysis. The process ensures the accuracy of data gathering, the reliability and validity of data analysis. Furthermore the auditor questions to what extent the steps undertaken in the study are appropriate, make sense and are acceptable in relation to the methodological standards of the domains the study was conducted in.

The process involves a dialogical interaction between the writers and an auditor starting with an audit contract between auditor and auditee where the timeline is established, the role of the auditor and the goals of the audit process specified. This step was done between the data gathering and the data processing, not correctly following the Akkerman et. Al. schedule but as the auditor is the first supervisor of the auditee she was already informed about the data gathering process. The steps of the data gathering process were also part of the annual supervisory agreement.

To give the auditor an orientation over the planned research steps a ‘ReadmeFirst document’ was created. It contains the research question, the conceptual framework the planned and used methods, the research setting including the description of the participants, the technical research tools, the Games selected and the expected results. The auditee prepared an audit trail in which the whole procedure of data gathering and analysis is documented in a process document, the documentation includes all raw data material as the transcripts of the semi-structured interviews, categorized and processed data in various steps, research tools like the interview guidelines, codes e.g. and the findings using a shared Dropbox between the auditee and the auditor. So the latter could access all data and follow every step more or less simultaneously and see what was done when, how and why.

Data Collection was clearly documented, missing data was accurately explained.

During the audit process the auditor could ask for the specification of unclear sections in the research documentation to guarantee the comprehensibility and acceptability of the study.

The findings are understandable and accurately documented.

The final document is not part of this audit process as it will be part of the dissertation it seems not appropriate that the auditor carries out two different evaluation processes at the same time with different goals and in different roles.

Result

The decisions made by the researcher during this process were visible, well argued, and acceptable considering the rules of reliability and validity in research in the Learning Sciences and Game Studies. The auditor concludes that the overall quality of this research is satisfying in terms of the reliability and validity.

Munich, 16th August 2016

Appendix F

MOBA jargon and concepts

Play roles	
Name	***Definition***
Support/Disablers	A support is a role in charge of making the teammates stronger, healing, shielding, providing boosts, and provide crowd control (i.e., stun, slow, snare, airborne, blind, disarm, root, suppression and silence).
Marksman	A marksman, also called an AD Carry (attack damage carry or physical attack) deals damage using ranged auto attacks, builds attack speed, and critical strike chance.
Midlaners	Midlaners do most of their damage via their abilities. Midlaners are champions that can deal high amounts of damage in a short time (bursts).
Tank/Initiator/Durable	A tank usually has high health and defence allowing them to safely initiate in team fights. Their function is to absorb damage, block skill shots, and distract the enemy.
Off Tank/Fighter/Bruiser/Carrier/Initiator	These three terms are sometimes used interchangeably. A fighter/off tank has some attributes of a tank (i.e., damage absorption, initiatiator, and uses used some crowd control –'CC'). A bruiser is a hybrid between tank and DPS (damage per second), combining the survivability of a tank and the damage of a caster (mage).
Jungler	The typical jungler is one who can survive fighting the jungle camps (clusters of jungle creep or monsters), with enough health to gank one of the lanes. Junglers usually have high health and a way to sustain their life, like life steal or spell vampirisation.

(continued)

Play roles (continued)

Name	*Definition*
Mage/Nuker	Mages, also called AP carry (ability point carry –magic attack) are mostly ranged champions prioritising abilities over basic attacks.
Assassin/Escape	Assassins specialise on infiltrating the enemy territory using stealth skills and quickly kill enemies via bursts.

Game phases

Name	*Definition*
Early Game	Also known as laning phase, starts when the game begins. At this phase players buy the basics items to start building their hero and survive the first battles. In this phase players have limited abilities. Early game ends when the first tower falls or when the first big gank happens. At this phase players' mobility is mostly restricted to the lane according to the role.
Mid Game	The mid game is dominated by ganks while targeting towers. At this phase players are not restricted to their lanes. This is where strategy starts to matter for the game. At this point other champions might start going to the jungle in order to take the buffs or farm gold and experience. During this phase, Dragon is an important objective and teams will try to take it. Although every game is different, a rule of thumb is that mid game ends when there are very few towers left.
Late Game	Late game is when all players are totally or close to totally developed in their experience, skills, and builds. At this point teams try to get Baron Nashor to get the buffs and try to take the main structures at the base of their opponents.

(continued)

Game phases (continued)

Name	*Definition*
Late Game	Because everyone is so powerful, teamwork, communication, and strategy are what matter the most to win the game.

Jargon

Name	*Definition*
Ability	One of the main unique powers a Champion can cast as a way to attack, defend, or support.
Autoattack	Basic attack by which a hero makes damage
Backdoor	A technique for which a player can directly access the enemy base without confronting the enemy and often taking advantage that the enemy is far from the base
Baron Nashor	The most powerful neutral monster in the Summoner's Rift in LoL. Killing it grants 600 experience to each team member, 800 experience shared between allies killing the monster, 300 gold for each member of the killing team, and a buff granting 40 bonus attack damage and ability power to each living member of the group.
Bot game	A game played with or against computer controlled heroes
Bot/Mid/Top	Short versions used by the payers to refer to bottom lane/middle lane/top lane of the map
Brush	Bush where the player can hide from the sight of other players or that blocks the vision of the enemy.
Buff	A positive temporal effect applied to a champion
Build/building a hero	The configuration of different items granting the player's hero certain abilities and strengths.
Camps	Clusters of neutral monsters that can be found in the Jungle.

(continued)

Jargon (continued)

Name	*Definition*
Champion/Hero	A particular character that the player controls. Champion is the term used in LoL and Hero the term used in DotA2
Debuff	A negative temporal effect applied to a champion.
DotA2	Acronym for the game Defense of the Ancients 2
Dragon/drake	It is the second most powerful neutral monster in the summoner's rift in LoL. Killing it grants 25 gold and 300 points of experience to the killer.
Farming	Seek and kill minions to obtain gold and experience.
Feed	Get killed by the enemy repeatedly, thus granting them money and experience.
Fog of War (FoW)	Dark areas on the map and the minimap that are out of the vision of the champions, minions, or structures.
Gank	To ambush an enemy with two or more champions.
GG (good game)	Message of sportsmanship indicating that either losing or winning the game, the player had fun. In some contexts it can be seen as a message that the player surrenders the game.
Inhibitor/Barracks	Structure preventing the contrary team to spawn super minions/super creeps. Inhibitor is the term used in LoL, and Barracks is the term used in DotA2
Lane	One of the main roads of the map that players and minions follow. There are three main lanes: bottom lane, middle lane, and top lane.
Last hit	To hit an enemy or a minion so that the enemy dies. Last hitting grants the player gold and experience.
LoL	Acronym for the game League of Legends

(continued)

Jargon (continued)

Name	*Definition*
Main	Player's favourite character or the character the player uses the most.
Mana	It is the pool of magic resource used for players to cast abilities.
Map awareness	To be conscious of the events occurring around the map.
Map control	To have vision and influence over areas around the map
Meta/metagame	Stage previous to the start of the game consisting on team and lane strategy set up.
	Also, the game outside the game. Gaming community outside the game dedicated to analyse the games in more detail (e.g., analysis of mathematical formulas behind character build or expansion of the game lore)
Minimap	Small version of the map at the corner of the screen granting the player an overview of what is happening in the map and game.
Minions/creeps/lane creeps	Computer-controlled unit helping the player to open their way to the other side of the map. They march in waves to the other side of the map following a designated lane.
Monster/neutral creeps/jungle creeps	Are monsters gather in camps in the jungle. Monsters do not attack unless attacked. Killing them grant gold and experience to the killer. Some monsters grant buffs to the players.
Nexus/Ancient	It is the main structure the teams must protect or destroy (in the case of the enemy team) in order to win the game.
Pings	A way of communication consisting of a noise and a signal that all the team can see in the map and minimap. Different games have diverse pings for communicating.

(continued)

Jargon (continued)

Name	*Definition*
Ranked game	A ranked game is a draft mode game available to players who have reached level 30 in LoL.
Ranked player	In LoL, a ranked player is a player that has reached level 30 and can play according to the ranking system starting in Bronze level. In DotA2 a ranked player must be at least level 20 and play 10 trial games to calibrate their Matchmaking Rating.
Roshan	It is the most powerful neutral creep in DotA2. Killing it grants 200 gold to each of the players of the killer team and 600 gold for the player last hitting it. It also drops the item 'Aegis of the immortals' which allows the player to automatically respawn upon dead.
Rune	In LoL a rune is a type of match boost part of the character build in the meta game. In DotA2 a rune is a type of buff that spawns in the river crossing the map.
Shop	Where players buy items to build their Champion or consumables such as potions and wards.
Skill	Word used as synonym of character ability
Spawn	When a monster, creep, or champion appears or is produced in the map
Spawn times	Time taken for creeps, minions, and monsters to be produced in the map.
Spell	Special skills used by the champions that can be chosen in the metagame. A character can choose up to two spells before starting the match.
Starter/initiator	A champion that performs an action which signals other players to start a battle.
Stats	The mathematical value of the champion skills (e.g., attack damage, armour penetration, lifesteal, etc.).

(continued)

Jargon (continued)

Name	*Definition*
Summoner	In LoL summoner is the name given during the game to the player controlling a designated champion. In DotA2 summoner makes reference to a hero who is able to spawn minions to help them in combat.
Super minion/super creeps	Special minions/creeps which spawn after the inhibitors/barracks have been destroyed. These minions/creeps have more health and do more damage than the average minion.
Tower/turret	Are the defending structures present in the lanes and which mission is preventing the players to advance close to the Nexus/Ancient.
Ulti/ultimate ability	It is the last ability that players can start developing in their characters as well as the strongest ability they can use.
Wards	Items that dissipate the FoW once they have been placed.
Lanes	
Name	*Definition*
Bottom Lane	It is the lane at the bottom of the map. Typically played by the roles of Marksman (Ranged Attack Damage Carry –physical attack) and Support.
Mid lane	It is the lane in the middle of the map. It is the shortest lane so the distance between turrets is also shorter. This lane also cumulates minions faster than the other lanes. Typically the roles Ranged AP, Assassin, or Melee AP
Solo Top Lane	It is the lane in the upper part of the map. Typically is played by only one player allowing for the jungler to make ganks. This lane is typically played by the tank, fighter, or off-tank

(continued)

Lanes (continued)

Name	*Definition*
Jungle	Jungle is everything in the map that is located between the lanes. Jungle spawns 'jungle creeps' or 'monsters' are found in the jungle in clusters called camps. Typically this position is played by the, jungler, tank, assassin, or fighter