2018

The Tracer Method: Don't Blink or You Might Miss it. A Novel Methodology Combining Cognitive Task Analysis and Eye Tracking

Kaitlyn Marie Roose
Michigan Technological University, kmroose@mtu.edu

THE TRACER METHOD: DON'T BLINK OR YOU MIGHT MISS IT. A NOVEL METHODOLOGY COMBINING COGNITIVE TASK ANALYSIS AND EYE TRACKING

By

Kaitlyn M. Roose

A THESIS

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In Applied Cognitive Science and Human Factors

MICHIGAN TECHNOLOGICAL UNIVERSITY

2018

This thesis has been approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE in Applied Cognitive Science and Human Factors.

Department of Cognitive and Learning Sciences

Thesis Advisor: *Dr. Elizabeth Veinott*

Committee Member: *Dr. Shane Mueller*

Committee Member: *Dr. Charles Wallace*

Department Chair: *Dr. Susan Amato-Henderson*

Table of Contents

List of figures

List of tables

Acknowledgements

I would like to, first and foremost, thank my advisor and mentor, Dr. Elizabeth Veinott (Beth). She has gone above and beyond to ensure that I am developing academically, professionally, and personally. She is one of the major reasons that I am able to complete my degree, and I am fortunate to continue to work with her for my doctorate. Thank you for all of the hours, the meetings, the feedback, and all of the support. Thank you for allowing me to study my two loves: cognitive psychology and video games. Thank you for "leveling" with me and pushing me to succeed. Thank you for telling me to stop saying "I'm sorry," and for being a strong female role-model in my life. Finally, thank you for not giving up on me, and thank you for bringing me to Michigan Tech.

I would like to thank my committee, made up of Dr. Elizabeth Veinott, Dr. Shane Mueller, and Dr. Charles Wallace. Thank you for reading through my proposal and providing me with valuable feedback. Thank you for supporting my research and encouraging me to dig deeper. My paper would not have reached this caliber without your support and critiques. Shane, thank you for asking me (twice) what the name of the method would be… I almost forgot.

David, even though you are a Mei main, I still love you. Thank you for all of the support and encouragement that you've given me, even from across the country. Thank you for never letting me give up, and for calling me every day. Also, I promised I would give you credit for the name of the method AND the tagline…Because after all, it was genius.

I could not have made it through a year of doing *Overwatch* research without the support of my eSports team. You guys not only stayed on the payload, but you also protected your healer (most of the time). Milk (TeHMILKDUD), thank you especially for the several hours that you put in to help me code my data. Thank you for being available and helpful, even at 2am. I'm going to miss you on the team.

Thank you Caden for your magic coding skills and helping me process and analyze the data. I'm glad that you're my "little," and I'm still super impressed at how you manage all of your commitments. Thank you for taking so much time out of your busy schedule to help your "big" out. I'm glad you aren't graduating yet.

I do not have the room to thank the many others that have helped me succeed thus far in my graduate degree. Regardless of the extent of your contribution or support, thank you.

Finally, I would like to thank my family, especially my father, mother, and brother, who have been my rock and the source of my homesickness. I miss you guys every day and I am so blessed to have your continuous love and support. Thank you for believing in me and supporting me throughout my life and for instilling in me strong values, and teaching me how to live with character. I thank God every day for my incredible family, and I cannot thank you enough for all that you have given me. Specifically to my brother, Gerardie: Thank you for being my first Player 2. I love you guys.

List of abbreviations

AOI: Area of Interest

CD: Critical Decision

CDM: Critical Decision Method

CIT: Critical Incident Technique

CoA: Course of Action

CTA: Cognitive Task Analysis

DPS: Damage Per Second

ET: Eye Tracking

FPS: First Person Shooter

Hz: Hertz

I-VT: Velocity-Threshold Identification

MMORPG: Massively Multiplayer Online Role Playing Game

MOBA: Multiplayer Online Battle Arena

OBS: Open Broadcaster Software

OW: *Overwatch*

PVAL: Primary Visual Attention Lobe

RPG: Role-Playing Game

RTS: Real-Time Strategy

SA: Situational Awareness

SME: Subject Matter Expert

Abstract

This thesis describes the development and first demonstration of a new Human Factors method, The Tracer Method, which is a combination of Cognitive Task Analysis (CTA) and Eye Tracking. The study evaluated whether the two methods together produce new and different information than either method alone could provide. The method was tested using a video game, *Overwatch*, a dynamic, complex, and multiplayer game. The evaluation included: 1. Examining both in the same context (game), 2. Establishing unique contributions of each method alone, and 3. Evaluating overlapping information. Results identified some overlap between the two methods that provided some cross-validation of the data. Cognitive Task Analysis provided higher level strategies and course of actions that players implement during their games, while eye tracking provided visual patterns of search (order of eye movements). However, when combined, the two methods provide strategy information in context that neither method alone can provide. CTA elicits insight into how individuals make decisions and apply previous knowledge, experience, and environmental information. Eye tracking can support this through predictive models of individual's eye tracking, to understand which elements are utilized in making predictions and situational assessments. We provide a tutorial and insight into best practices for implementation of The Tracer Method. This is the initial development of the new method, and on-going research is validating it in different environments. The Tracer Method is the first combined and documented systematic methodology that utilizes a changing and complicated environment and tests the interaction and output of Critical Decision Method and Eye Tracking.

Keywords: Cognitive Task Analysis, Eye Tracking, Critical Decision Method, Video Games, Human Factors, Knowledge Elicitation, Decision Making

Chapter 1: Introduction

Video games have taken the world by storm, with many children averaging 10,000 hours of gameplay by the time they reach 21 years old (McGonigal, 2011). Video games vary across genre, style, strategies, and goals, but in all video games, players are challenged to make decisions. Developers and researchers are intrigued by what happens during gameplay, and need methods to systematically understand and evaluate game play at various stages of development from playable to beta versions. Decision making in games is both interesting and important to researchers and designers. Understanding how people solve problems and make decisions is key to improving decision making and developing impactful games (Roose & Veinott, 2017; Veinott et al., 2013; 2014). Decision making isn't just important in games. Marketing, politics, military, athletics, and medical fields seeks to understand and utilize experts' decision making strategies to better inform training and system design. How do we elicit information from experts? How can we pick their brains that are filled with years of experience?

Cognitive Task Analysis (CTA) methods have been utilized to unpack the cognitively complex activities in many dynamic environments like video games, and can be used to assess the efficiency of mechanics, rules, challenges, and other game design elements within video games. Past research using CTA in video games has been limited to only using a few of the many CTA methods (i.e. verbal protocols) and focusing on usability or transfer of learning. Researchers have not yet unlocked the full potential of the CTA toolkit. This thesis describes the study used to test and evaluate The Tracer Method in the context of Overwatch, a complex multiplayer first person shooter (FPS) that has exploded in popularity and reaching 35 million users within 1.5 years of its release.

To understand the cognitive, strategic, communication, and perceptual aspects of cognition that are present in the Overwatch FPS game, I developed The Tracer Method. This method combines one of the most studied and replicated CTA Methods (Critical Decision Method) with a commonly utilized and commercially available method, Eye Tracking (Figure 1). The Tracer Method demonstrates the compatibility of these two methodologies and how CTA provides the necessary context for the behavioral data from ET and guides the researcher to information that is most critical to understand and unpack. This research both builds of off the past, but also documents a systematic novel methodology for unpacking decision making in complex environments. One contribution this paper is making to game research is demonstrating how CTA can be used to unpack critical cognitive activities in games. To the extent that this combination is effective, this contribution will not only be important for game researchers, but also for human-computer interaction researchers.

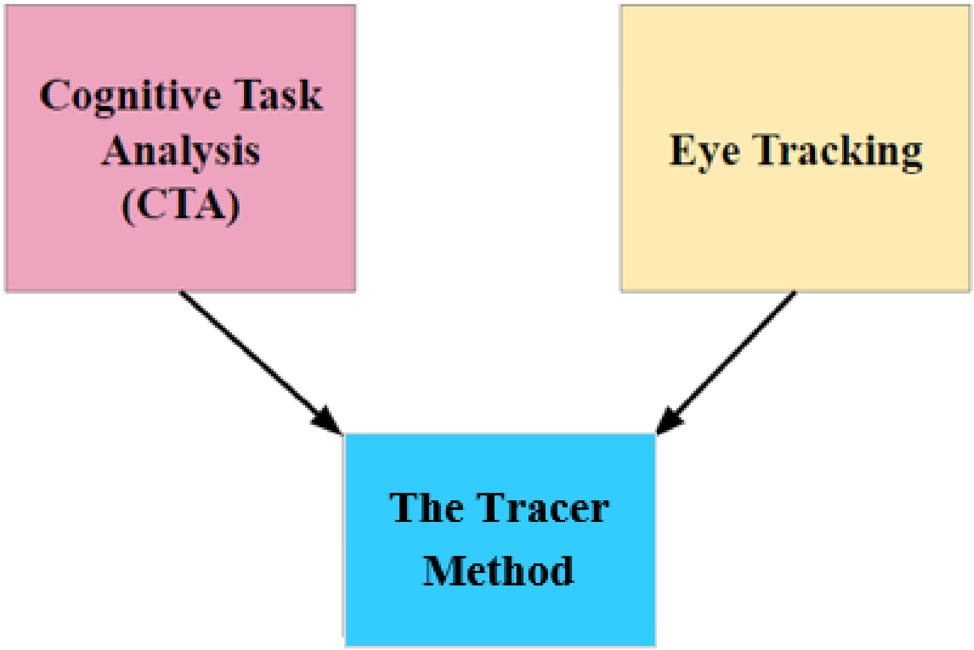

Figure 1 The Tracer Method, formed by combining Cognitive Task Analysis and Eye Tracking methodologies into one.

Because the Tracer method combines two widely-used methods, I will first review the background CTA and ET in general, and how the methods are applied across domains and if they are utilized within video games. I will then discuss the limited research that utilizes the combination of the two methodologies.

1.1 Cognitive Task Analysis (CTA)

Cognitive Task Analysis is a set of methods used with experts to unpack their cognitive processes during complex decision making. CTA methods have been developed, utilized, and validated over the course of several decades (Crandall, Klein, & Hoffman, 2006), in a variety of environments: military (Kaemph et al., 1992; Klinger et al., 1993; Schaafstal et al., 2000), aviation (Brezovic, Klein & Thordsen, 1991; Hutton et al., 1997), medical (Crandall & Calderwood, 1989; Crandall & Gamblian, 1991; Crandall

& Getchell-Reiter, 1993), and civil (Calderwood et al., 1987; Crandall, 1989). Some common methodologies include: Critical Decision Method (Hoffman, 1987; Hoffman et al., 1998; Klein et al., 1989), structured interviews (Crandall et al., 1994; Hoffman et al., 1995; Randel, Pugh, & Reed, 1996), Hierarchical (Annett, 1996; Stanton, 2006) and Goal-Directed Task Analysis (Endsley et al., 2003), and simulation interviews (Mislevy et al., 1999).

Historically, most CTA methods are based on Critical Incident Technique (Flanagan, 1954), which consists of a series of methods to observe and gather information about human behavior in concrete situations. This technique often provides guidelines to decrease errors and increase performance for training novices. CTA also shares its roots with Task Analysis (Cooke, 1994; Jonassen, Tessmer, & Hannum, 1999), which focuses on the tasks that require management of complex cognitive resources and strategies. This technique allows the researcher to understand the hierarchy of the tasks in addition to the behavior implications of the tasks.

This technique often yields guidelines to decrease errors and increase performance. CTA also shares its roots with Task Analysis (Cooke, 1994; Jonassen, Tessmer, & Hannum, 1999), which focuses on the tasks that require management of complex cognitive resources and strategies. This technique allows the researcher to understand the hierarchy of the tasks in addition to the behavior implications of the tasks. Cognitive Task Analysis has been developed and refined for over 50 years. It evolved from task analysis to cognitive task analysis with the growth of human factors psychology (Hoffman & Militello, 2012). From CTA, mathematical and computer models have been developed to understand the interaction of humans and technology in order to improve system and algorithm design (Schweickert et al., 2003). CTA is used to focus on job requirements and performance, however, cognitive psychologists have pushed for the additional cognitive component to be added. Cognitive Task Analysis isn't simply an interview method, but is rather systematic and purposeful (Crandall et al., 2006; Ericsson & Simon, 1980; Nisbett & DeCamp Wilson, 1977). Cooke (1994) articulates the benefits of a semi-structured interview (such as CDM): 1. Heavy focus on goals, 2. Complete coverage of field experience, 3. Gaps in knowledge are more easily addressed, and 4. Knowledge elicited is more focused on critical elements or tasks versus general procedures. For many of the methodologies, there are interview guides or prompts that have been used in many studies. Our study uses probes directly from Crandall et al. (2006).

Furthermore, Cognitive Task Analysis can consist of the utilization of one method or multiple methods (Crandall et al., 2006). Multiple methods allow for distinct and numerous types of information to be used for both collecting, analyzing, and understanding the data. Typically, CTA is done with experts in a domain, field, or job in order to elicit knowledge structures and create models for behavior (Schraagen et al., 2000). CTA elicits knowledge and focuses on qualitative data (Militello & Hutton, 1998). CTA data can result into the development of design requirements or changes, training procedures or scenarios, or measures from experiments. Combining a qualitative methodology such as CTA with a quantitative methodology (Eye tracking) could provide new insights and perspectives on cognitive perspectives that either one cannot deliver alone.

1.1.1 CTA in Video Games

A few studies have attempted to use Cognitive Task Analysis techniques to develop design requirements (Boyle et al., 2012; Gallagher and Prestwich, 2013) or better understand game play for learning (Horn et al., 2017; Shute & Kim, 2011). For example, Gallagher and Prestwich (2013) concurrent verbal think aloud and retrospective cognitive mapping to examine the effect of design features on player navigation in Portal 2. Horn et al., 2017 interviewed players to identify novice skill chains (e.g., triple jump in Super Mario 64). Their methodology is not typically considered a CTA. Shute and Kim (2011) analyzed verbal protocol data to understand learning strategies in a physics-based puzzle game. Results indicated participants did not feel that the game translated to their academic studies. The concurrent verbal protocol articulated their engagement and how they answered the prompts during the gameplay. The participants created mappings of in game concepts to classroom knowledge, which was found to be inconsistent and difficult to articulate. Boyle and colleagues (2012) used a semi-structured interview technique to elicit knowledge from teachers in research methods and statistics. They developed guidelines for developing a video game to help support students in learning these subjects. These studies have used both process tracing methods and semi-structured interview techniques to elicit knowledge, strategies, and input from individuals playing games or with the intent to analyze game design. CTA has many useful methodologies to support this elicitation and utilization of information involving video games, however, the CTA toolset has not been extensively utilized.

1.2 Eye Tracking (ET)

Eye tracking (ET) records and measures eye movements when the eyes are looking at, following, or interacting with an element in an environment. Eye tracking helps us understand specific eye movements, primarily fixations and saccades (Duchowski, 2007; Poole and Ball, 2006; Salvucci and Goldberg, 2000). Measurements of fixations include: gaze durations, gaze patterns, and areas of interest (AOIs) analyses. Saccades are measured through saccadic velocities, saccadic amplitudes, and AOIs. Eye tracking has been used in usability studies (Goldberg and Wichansky, 2003; Jacob and Karn, 2003; Poole and Ball, 2006), learning (Judd et al., 2009; Rehder and Hoffman, 2005), and attention research (Granka et al., 2004; Keefe and Silverman; 1997). The goal is to understand visual strategies during various tasks.

Historically, eye trackers have been around since the 1800s, with the earlier technologies restricting head movement (Huey, 1908; Holmqvist et al, 2011). Nivvedan (2014) reviewed the earliest studies investigated eye movements in text reading (Buswell, 1937; Dearborn, 1906; Just & Carpenter, 1980). They described that eye movements are not as smooth and previously assumed, but are made up of short and rapid eye movements (fixations and saccades). Specifically, Just and Carpenter (1980) developed a reading model that the fixations are longer for higher processing loads. The points of greater load tended to occur when readers stumble upon unfamiliar words or phrases, integrating critical information, and making predictions.

Eye tracking technology has improved dramatically by decreasing the need to restrict the head, and increasing mobility. The way we understand the visual system, human attention, and cognition has impacted the development of new technology (Duchowski, 2002; 2007). Current studies of eye tracking investigate attention in automotive (Land & Lee, 1994; Recarte & Nunes, 2000), medical (Law et al., 2004; Mello-Thoms et al., 2002), aviation (Anders, 2001; Ottati, Hickox, & Richter, 1999), and video game environments (Alkan & Cagiltay, 2007; Almeida et al., 2011; El-Nasr & Yan, 2006; Jonsson, 2005; Kenny et al., 2005; Leyba & Malcom, 2004; Renshaw et al., 2009). From eye tracking data, we can further understand people's navigation of the environment, how they gain information visually, and some insight into visual strategies implemented during complex tasks. For example, Law et al. (2004) concluded that experts and novices had different eye movements and strategies in virtual aiming tasks. Novices tended to look at different tools for longer periods of time, and required more visual feedback from the tool positions. Experts, on the other hand, were able to maintain their gaze on the target while not having to continuously gaze at the tool position.

Eye tracking is typically used in less dynamic environments in order to evaluate usability. Duchowski (2002; 2007) describes how eye tracking output can provide diagnostics on elements and interactions of various stimuli. For example, a measure of efficiency may be number of fixations (Poole & Ball, 2006). The more fixations an individual has to complete a task, the more inefficient the search or visual strategy is. In visual search, a similar measure can be used in addition to fixation duration, which can articulate the complexity and needed processing time of different elements of the user interface. Individuals may require more guidance in order to navigate the environment. For example, Duchowski (2002) describes usability studies in the cockpit with the integration of new technology such as electronic map displays. The difference between eye tracking measures (such as fixation durations or number of fixations) before and after the integration can articulate performance and efficiency differences.

1.2.1 ET in Video Games

Eye tracking has been used as both an assessment methodology and as an input for controlling the game environment (Almeida et al., 2011). Video game researchers have used eye-movements to examine player strategies in games (Alkan & Cagiltay, 2007; Almedia, 2009; Kaufman et al., 1993; Kennedy et al., 2005; Renshaw et al., 2009). El-Nasr and Yan, (2006) investigated visual attention in two types of 3D video games: action adventure and first person shooter (FPS). Their evidence suggests that in action adventure games, players are using a top down approach more often than a bottom up approach. Typically, in first person shooters, they found that the eye movement patterns hovered over the middle of the screen with the reticle, rather than a larger range of eye movements in action adventure games. Almedia, (2009) examined how eye tracking could be used to develop usability guidelines for video games. Their findings justifies the positives for using eye tracking to evaluate games and they arrived at a similar conclusion: eye tracking provides new insights into gameplay and usability in games.

When combined with a method such as retrospective think aloud protocol, eye tracking improves and validates information pulled from players.

Kenny and colleagues (2005) examined whether eye tracking to could improve rendering and develop better algorithms for a First Person Shooter (FPS) game. They developed a system that co-registers eye movement information in combination with the virtual environment presented to them. They also suggest a potential correlation between task proficiency in games and fixation and movement order durations. Specifically, with players having high percentages of fixations in the center of the screen (and longer fixations) correlating to higher scores. Alkan and Cagiltay (2007) used eye tracking to unpack students' learning in a game. They found that a trial-and-error strategy (from both interview and ET) was most common among participants. Similar to the early eye tracking studies (Just & Carpenter, 1980), they found that participants' fixation duration was the longest in the "contraptions area" which is considered the place where people evaluate and create potential solutions. Eye tracking can support the identification of problem solving strategies, especially in simple settings. Renshaw et al., (2009) used ET with 7 players to examine different ET patterns under different time pressure situations during a puzzle game. Although their research focus was on engagement, they concluded that when players were not performing a specific task (i.e. moving or shooting at something), they fixations were more spread across the screen rather than in the middle of the screen. While this study has some interesting ideas, it did not have the statistical power to assess the hypotheses. In CTA studies, there are not typically as many participants as ET studies. This study only investigated 7 people, and the authors identified their findings as preliminary.

Other studies have examined using eye movements as input (Leyba & Malcom, 2004; Smith & Graham, 2006). Several game researchers suggest that ET is useful for assessing navigation (Kaufman et al., 1993), player interaction (Jonsson, 2005), and usability (Almedia, 2009; Johansen et al., 2008). Using eye tracking, Almedia (2009) was able to identify map locations that participants were using and referencing and how they impacted their gameplay. Smith and Graham (2006) explored the use of an eye tracker as an input control device in several types of games: FPS, Role-Playing game (RPG), and Action. Players feel more immersed in the game environment when their eyes are controlling elements of the game (i.e. view or avatars). Players also described a decrease in workload, as it reduces motor movement that attaches to navigating the environment, so they can focus on other elements. Leyba and Malcom (2004) compares eye tracker aiming versus mouse aiming in a computer game. They found that participants were more accurate using the mouse than their eye inputs, however, this was attributed to the need for a better calibration process with the eye tracker.

Eye tracking is an efficient method to analyze visual assessment patterns of individuals as they navigate through dynamic (or static) environments. However, ET may not provide all contextual elements of the data, thus, leaving a gap in the interpretation. Combining ET with other methodologies may allow researchers to close that gap and view their data through a different lens.

1.3 CTA and ET in Combination

Very few studies are at the intersection of CTA and ET. Most of the literature involves the analysis of eye movements compared to what people say during concurrent or retrospective think aloud protocols (Elling et al., 2011; Elling et al., 2012; Guan et al., 2006; Kaakinen & Hyona, 2005; Olsen et al., 2010; van Gog et al., 2005; van Gog et al., 2005). These studies demonstrate a direct connection between what people are looking at and what they are saying in the think aloud protocols. This data provides some construct validity between data from the verbal protocols and the eye tracking movements. Other research has investigated eye tracking as a supplementary methodology for general process tracing (Gidlof et al., 2013; Glaholt & Reingold, 2011; Renkewitz & Jahn, 2012). Seagull and Xiao (1999; 2001) used a previous task analysis from experts for a complex medical procedure as the basis for an eye tracking assessment that occurred during the same medical procedure. They asked SMEs to address what perceptual cues were used during this medical procedure and to evaluate video footage of participants performing the task. The authors' goal was to ask the SMEs to make predictions and justify why participants exhibited various eye tracking patterns.

Studies have mainly focused on combining the verbal protocol and process tracing methodologies with eye tracking, which has been effective. The data from these methodologies is directly mappable to the dynamic processes that are articulated from ET data, which is a large benefit to using verbal protocols. For most of the research done using verbal protocols and process tracing methods with eye tracking have focused heavily on calculating the amount of overlapping information. Cooke (1994) articulates that these methodologies are unable to capture all of the knowledge associated with completing a task or making a decision. However, to date, there has been no research investigating any other Cognitive Task Analysis method (i.e. Critical Decision Method) and eye tracking, thus, this would be the first study to do so.

2 Chapter 2: The Tracer Method

2.1 Motivation

We decided to begin developing this novel methodology for several main reasons: technology is advancing and humans are still cognitively engaging in these new advancements. Researchers, designers, academics, and industry leaders need effective and easy to implement methodologies that allow them to rapidly gain actionable and meaningful feedback. Within the human factors discipline, the combination of CTA and ET has never been evaluated outside of a simplistic and static environment or task. Because of the difficult, complex, and engaging environments that “real” people are presented with every day, the methodologies used to unpack cognition during decision making or other mental activities need to remain scalable and universally applicable across domains and varying complexities.

Previous research in usability has found benefit in the combination of process tracing methods and eye tracking. The Tracer Methodology takes advantage of an individual’s cognitive processes relating to decision making to both identify and leverage the critical points of interaction or execution. Figure 2 depicts a 30-60 minute task and where both The Tracer Method, and previous methodologies would sample, analyze, and apply the data.

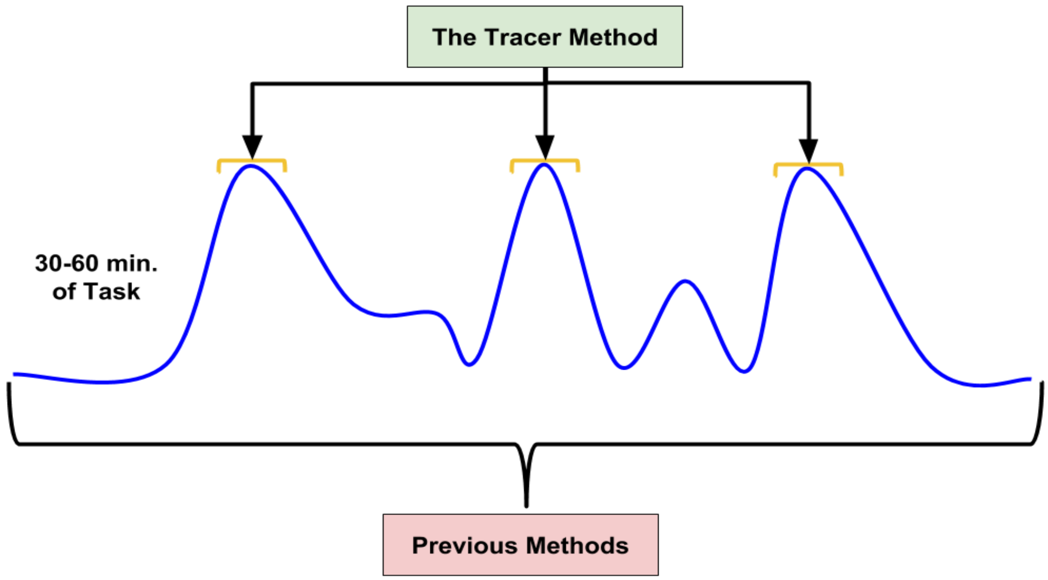

Figure 2 A visual depiction of a 30-60 minute task, and how both methodologies approach data collection, analysis, and application.

Because The Tracer Method does not focus on "generalizations" or strict procedural behaviors, it provides actionable and detailed feedback relating to specific elements of the design or interaction. The Tracer Method is flexible in allowing the researcher to choose a methodology that most closely helps them answer their research questions and gain the most meaningful output from the data collection and analyses.

2.2 Definition

Although The Tracer Method is still in its infancy, there are defining characteristics of this developing methodology that separate it from previous CTA + ET combinations. 1: The CTA methodology chosen must not focus strictly on procedure, but cognitive elements or behaviors associated with that procedure or execution. 2: The task or environment must be both cognitively and "physically" engaging and challenging (at some point). Meaning, playing a game of Sudoku would not require the use of this method due to its lack of graphical/physical challenge or depth of processing. Similarly, The Tracer Method would not be used in situations in which the individual is not strictly engaged (e.g. watching a movie). 3: The Tracer Method must be used for multidimensional or ill-defined problems or environments. Situations in which there is a singular or linear solution would be an inefficient use of time, and would not provide variability in responses. Overall, The Tracer Method would be defined as: A systematic methodology combining CTA and ET in order to analyze cognitive and behavioral elements of an engaging, challenging, and non-linear environment or execution task.

2.2.1 Which CTA Methods Constitute *The Tracer Method?*

CTA has many methodologies within its toolbox, so do all of them count as The Tracer Method when combined with ET? The short answer is: No. Specifically, which CTA methods are considered part of this methodology. As this is the first demonstration working to develop this method, it is impossible to specify with 100% certainty which methods are acceptable. However, process tracing methods, such as think-aloud protocol provide a good starting point. Although both concurrent and retrospective think alouds have their benefits and applications, they provide limited insight and "force" individuals to describe behaviors or engagements with a technology or task that may or may not have more depth to them. Similar CTA methods that scratch the surface would be inefficient to use, and would be more beneficial using them in a separate context from ET. These methods are also very heavily focused on procedure, restricting how the data can be categorized or grouped in the analysis stage. There is little to no deep cognitive insight into their thought process or other cognitive behaviors (e.g. planning, decision making, or problem detection). The output is more focused on both procedure and outcomes that are conditioned on either the accuracy of these procedural steps. Methods such as Simulation Interview or Knowledge Audits provide a specific and useful structure that is still flexible enough to conform to your research question, objective, or project environment. There is still much research to be done to validate, replicate, and further refine The Tracer Method, however, the first demonstration in a fast paced, complex environment of a first person shooter provided a significant initial challenge to which The Tracer Method rose to the occasion.

3 Chapter 3: Research Questions and Hypotheses

3.1 Research Questions and Hypotheses

This paper addresses three types of research questions: methodology related (MR), decision making related (DM), and *Overwatch* related (OW). Methodology research questions investigate the extent to which The Tracer Method is useful as a combined methodology. We will use measures of overlap consistent with previous research, in addition to individually unpacking the output from each method separately to see what new information can be elicited. Decision making research questions examine the extent to which The Tracer Method further explains decision making in this context. We will use the output of each separate method and the combination to describe what we learned about decision making. Finally, *Overwatch* related research questions are specific to the information elicited using this methodology within the context of the game. There has been no research executed using *Overwatch* as the testing environment, and this research would be the first to unpack what types of decisions are being made in this game, what type of information are players using to make decisions, and what can we learn from *Overwatch* that can be applied to other domains. The methodology related (MR) research questions are central to this thesis and their findings will be described at length in the main results sections (Chapter 5). The other two (DM & OW) sets of research questions are supplemental, and will be discussed in a later section (Chapter 6).

3.1.1 MR1: To what extent does CTA provide different information than eye tracking?

Past research has used think-aloud protocols, which are tightly coupled to eye-tracking as people are reporting exactly what they are doing. Therefore, the amount of overlapping information is expected to be lower because of the difference between CDM and verbal protocol (Cooke, 2010; Elling et al. 2011; 2012), however, the exact percentage is difficult to hypothesize. Thus, the amount of different, or unique information will be higher than the amount of overlapping information.

3.1.2 MR2: To what extent does the combination of a CTA and ET provide new information that either method could not alone?

We are unable to hypothesize about the specific extent of new information that The Tracer Method would produce that either method could not separately provide. We are predicting, however, that this new methodology will provide new insight that we do not know the nature of. Because CDM has never been used with ET, this also complicates the prediction of the type or extent of new information.

3.1.3 DM1: To what extent do Strategic, Operational, and Tactical decisions differ in number of decision cues and courses of action?

Hypothesis 1: Tactical decisions will identify using the most cues compared to Operational and Strategic decisions. Strategic decisions will identify the most courses of action (CoA). Tactical decisions occur on a shorter temporal scale, and thus will require more information “immediately.” Strategic decisions will identify more courses of action than the other two decision types because Strategic decisions may involve predicting multiple outcomes due to its larger time scale and overarching goal orientation.

3.1.4 DM2: To what extent do Coordination, Situational Awareness, Sensemaking, and Managing Uncertainty and Risk decisions differ in number of decision cues and courses of action?

Hypothesis 2: Sensemaking and Managing Uncertainty and Risk decisions will identify more cues and courses of action than Situational Awareness and Coordination decisions.

3.1.5 DM3: To what extent do Strategic, Operational, and Tactical decisions differ in distribution of fixation and visit counts, and fixation durations across AOIs?

Hypothesis 3: Strategic decisions will have the most diverse distribution of fixation and visit counts across AOIs, and longer fixation durations across AOIs.

3.1.6 DM4: To what extent do Coordination, Situational Awareness, Sensemaking, and Managing Uncertainty and Risk decisions differ in distribution of fixation and visit counts, and fixation durations distributions across AOIs?

Hypothesis 4: Sensemaking and Managing Uncertainty and Risk decisions will have greater diversity in fixation counts across AOIs, and Sensemaking and Coordination decisions will have longer fixation durations across AOIs.

3.1.7 OW1: To what extent DPS, Tank, and Support players differ in number of decision cues and courses of action identified?

Hypothesis 1: Tank players will identify the most cues and courses of action. Likely due to their dual roles in both offensive and defensive tactics, Tanks are likely to elicit more cues and identify more courses of action.

3.1.8 OW2: To what extent do DPS, Tank, and Support players differ in number of Strategic, Operational, and Tactical decisions?

Hypothesis 2: Tank players will make the most Tactical and Operational decisions, and Support players will make the most Strategic decisions. As Support players are likely in the backline and have the best view of both teams, they will likely be making Strategic decisions. With Tank's dual roles as offensive and defensive entities, Tanks tend to engage in many short term decisions.

3.1.9 OW3: To what extent do DPS, Tank, and Support players differ in distribution of fixation and visit counts, and fixation durations across AOIs?

Hypothesis 3: We are unable to hypothesize about the specific differences across roles in eye tracking behaviors, as no prior research has suggested or investigated differences. We do predict that if there are differences across decision types in eye tracking behaviors, it may carry over by role, or that decision type may interact with role.

This research describes the development of a new methodology (The Tracer Method) and evaluates the contribution of each sub methodology. We seek to understand both the unique and overlapping information to determine the effectiveness and necessity of combining these sub methodologies into The Tracer Method. For this project I conducted CTA interviews with experienced *Overwatch* players, had them play a competitive *Overwatch* while recording their ET behavior. We believe that the whole (CTA + ET) is greater than the individual contribution of each methodology alone. While we could evaluate these methodological questions in a driving simulator, on an aviation interface, or with a local hiking group during route planning, we are testing it in *Overwatch*, a popular multiplayer FPS with over 35 million players within 1.5 years of release.

4 Chapter 4: Methods

4.1 Participants

Seventeen *Overwatch* players (94% male) with a mean age of 21 years (SD=3.74 years) were recruited from Introductory Psychology and multiple student organization pools. Students in Intro to Psych were the only participants that got course credit. The recruitment population was skewed for gender, with only a few females identifying as Overwatch players. This discrepancy was not by design. To participate in the study, participants had to meet several game-related criteria: 1. Reached at least level 25 in Overwatch, 2. Played at least 50 hours, and 3. Have experience specifically in the competitive mode. On average, participants have been playing Overwatch for 19 months. At the time of data collection, the game had been publically released for about 18-19 months. Participants play an average of 8.3 (SD=7.2) hours of Overwatch per week, with time fluctuating during the academic year. We did not ask participants what other games that they play or how much time they play other games, and wanted to focus specifically on Overwatch expertise. Most participants (87%) typically played with their friends. Across all participants, there was a relatively even distribution of main role identification: 6 Damage Per Second (DPS), 6 Tank, and 5 Support players. Only one member of each of the 6-man teams were interviewed.

4.2 Data Collection

Data for this study was collected through multiple CTA methods in addition to ET data. Two software packages were used for recording and data collection. Blizzard's

Overwatch (2016) was chosen as the game environment for the demonstration of The Tracer Method.

4.2.1 CTA: Task Diagram

The Task Diagram is a knowledge elicitation technique that seeks to understand cognitively difficult tasks. This methodology is useful for gaining a general understanding of the themes of the data, and how it could help shape future interviews (e.g. a CDM). Based on participant's best or most experienced role, they were asked: "As a [role], what are 3 of your cognitively difficult tasks during the game?" Then, for each cognitively difficult task, they were asked:

1. *Why is this cognitively difficult?*
2. *What are your strategies for managing [the task identified]?"*

The outcome of the Task Diagram allows us to understand the overarching tasks that are occurring in *Overwatch* and where the overlap and differences are among roles.

4.2.2 CTA: Critical Decision Method

The Critical Decision Method (CDM) was used in this study to systematically explore key decisions made during a competitive Overwatch game. CDM is structured interview technique that unpacks cues, strategies, and courses of action in a way that can be difficult to capture in questionnaires (Crandall et al., 2006; Klein et al,1989; Hoffman et al,1998). Interviews were conducted right after game play and focuses on identifying the critical decisions through a series of four interview sweeps. CDM involves an in depth interview that contains 4 steps: 1) Incident Identification, 2) Timeline Verification,

3) Deepening, and 4) “What If” Queries (Appendix A for interview guides). During the first sweep, the interviewer establishes an overview of the key or critical decisions. Then, the interviewer develops a timeline of all the critical decisions. The next sweep involves a deepening sweep where one asks a series of questions designed to get the story behind the decision. This sweep includes questions regarding which cues, goals, priorities and actions were used. During the final sweep, participants were asked about expert-novice differences in Overwatch play. For example, “what might a novice do in this situation?”

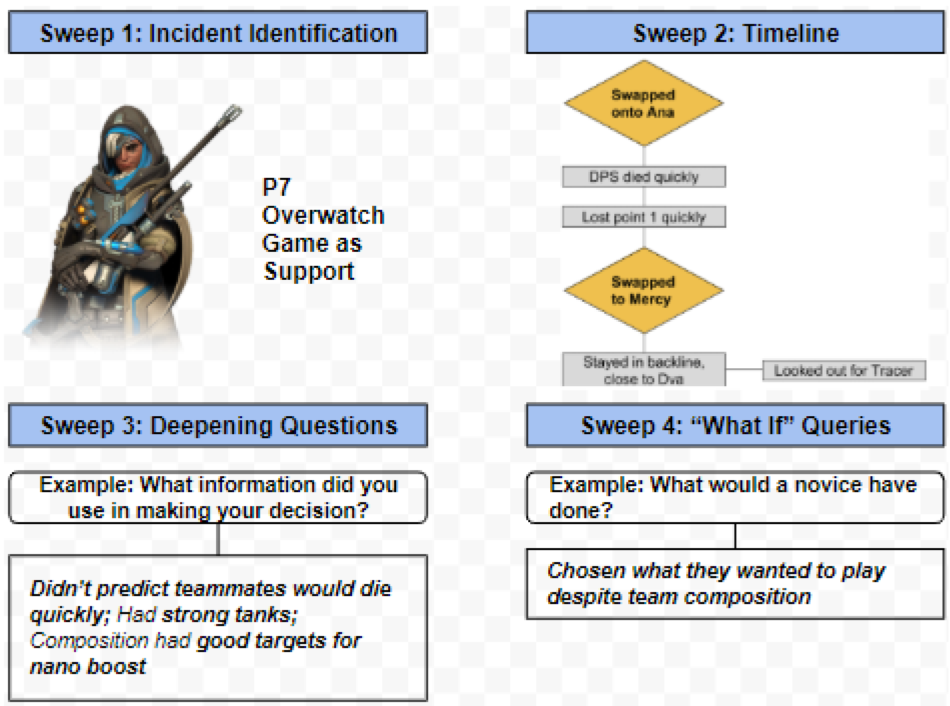

Figure 3 An example CDM walkthrough with excerpts from Participant 7's (Support) interview. Critical decisions are depicted as gold diamonds in the timeline.

Figure 3 depicts a CDM walkthrough with excerpts from the data that answer example probes for each sweep. The goal of the CDM is to systematically elicit decision information from experts by deeply unpacking the critical decisions. Critical Decision Method is one of many CTA methods that is useful for identifying expert critical decisions, cues, CoAs, and processes during complex incidents. Specifically, CDM is especially helpful in identifying expert-novice differences and breaking down expert processes in order to better inform training or develop models of behavior.

4.2.3 ET: Tobii Pro X3-120 Tracker and Tobii Pro Studio Software

The Tobii Pro X3-120 Eye Tracker was used to collect eye tracking data. The eye tracker was mounted at the middle of the bottom of the screen and was not intrusive to players while collecting eye movement data (Figure 4 below).

Figure 4 Monitor set up with mounted Tobii X3-120 Eye Tracker (outlined in gold).

The gaze sampling frequency is 120 Hz with 97% trackability, 0.24° precision, and 0.4° accuracy. The system latency is less than 11ms. Each participant was calibrated to the eye tracker before their movements were recorded, using a 9-part calibration. The eye tracker required that participants be at a distance from 40-80 cm away from the screen, which was determined to be good or not immediately after calibration. This distance was adjusted and was not standard across participants, however, all participants' distances and heights needed to be within the green (which represented good distance). Participants did not use a chin rest or anything that restricted their head movement. They were free to move their head (within reason).

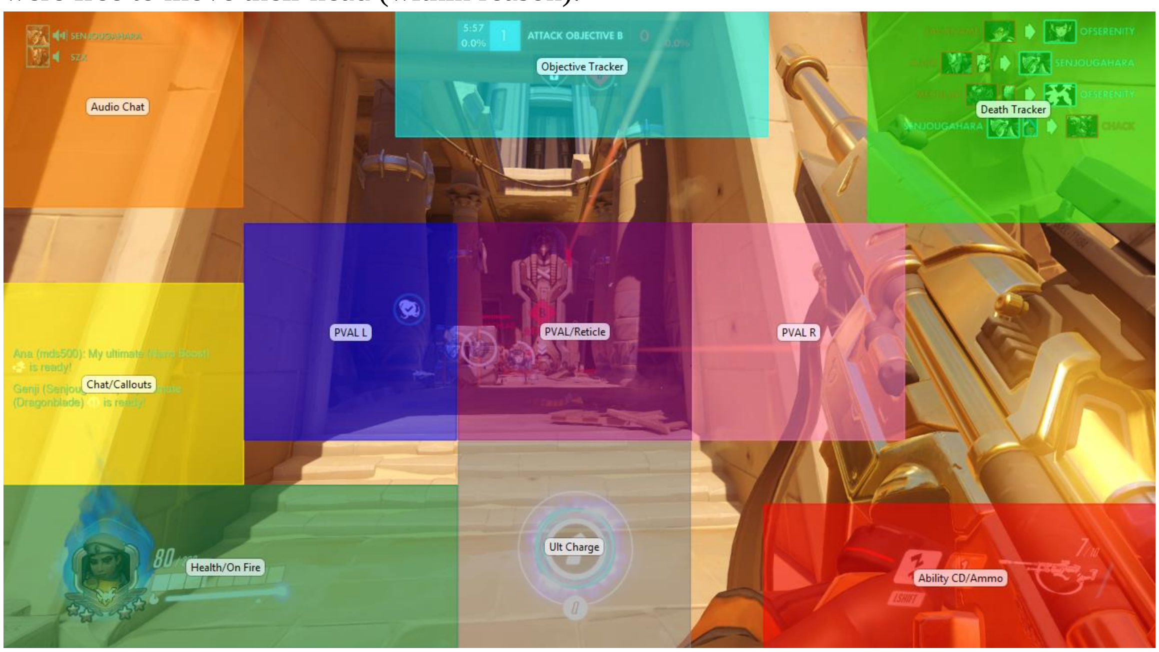

Figure 5 Screenshot from Tobii Pro Studio depicting the AOIs overlaying the Overwatch user interface.

To define fixations, we chose the default I-VT fixation filter in Tobii Pro Studio, which calculates eye movements based on the velocity of the directional shifts of the eye. From the eye tracking literature, a minimum fixation duration has not been defined, so we kept the minimum at the default of 60 ms. Within Tobii Pro Studio, the researchers created pre-existing AOIs based on the elements of the Overwatch user interface (Figure 5). These AOIs were used in the proceeding data analysis. Several visualization types, such as gaze plot (order of fixations) and heat map (gaze opacity) were also used for different analyses in this research. Tobii Pro Studio was used for the design, capture, and analysis.

4.2.4 Audio Recording: Open Broadcaster Software (OBS)

Open Broadcaster Software is an open source software for video recording and streaming. Audio recording was the primary feature used for this experiment. Gameplay was not necessary to record using this software.

4.2.5 Game: *Overwatch*

The Overwatch first-person shooter (Blizzard, 2016) was used as a testing platform. At the time of data collection, the game had only been out for 1.5 years and had over 35 million players. The Overwatch League (OWL) has launched as the game's pro league, and is in its first season. Players have the opportunity to participate in sponsored Overwatch tournaments either individually or as a part of their university (via Tespa). Overwatch has won many game awards, including 2016 Game of The Year and 2017 Best Ongoing Game.

4.2.5.1 Team Roles and Composition

In *Overwatch*, there are four role categorizations of heroes: offense, defense, tank, and support. Typically, players group heroes based on their "role" or responsibilities in games: DPS, Tank, and Support (typically healer). Heroes in the defense class are distributed into these “roles” based on the needs of the team and abilities of the individual heroes (e.g. Widowmaker is a sniper labeled a DPS). Each role usually has unique characteristics and goals, however, some characters possess ability types from multiple roles (e.g. Soldier 76 is a DPS that has the ability to place a healing field on the ground for teammates). In the context of this study, we refer to the three roles as: DPS, Tank, and support. These roles are often described as:

- **Damage Per Second (DPS):** Heroes that have a higher damage output and traditionally the most offensive potential. Their goal is to eliminate enemy heroes, especially enemy supports, and maim or put pressure on enemy Tanks.

- **Tank:** Heroes that usually have higher sustainability due to their large health pool and are able to absorb more damage and protect teammates from enemy fire. Tanks can be classified as “main” or “off-tank,” which is distinguished by their protective functionalities versus more offensive oriented abilities. Their goal is typically twofold: to both help fortify their teammates and disrupt enemy engagements.

- **Support:** Heroes that have the ability to heal teammates, augment their team’s abilities or sustainability, or debuff enemies. With the exception of one hero in the current support class, the term “healer” is conventionally attributed to the role. Their primary goal is to keep their teammates alive while escaping danger from the enemy (as they are usually a primary target).

Teams are made up of 6 members, and composition can consist of any combination of the roles (i.e. 2 DPS, 2 Tanks, 2 Supports). There are 296,010 possible combinations of team compositions that can be used by teams. Any player can play any of the 27 current heroes, however, there is a limit of one individual hero per team. Players can switch their hero (or role) at any point throughout the game, allowing teams to flexibly execute and adjust their plans as the need arises.

4.2.5.2 Game Modes and Objectives

Overwatch can be played in different modes from quick play to competitive games, there are two teams of six players with each team attempting to accomplish various objectives. There are four types of game modes that players may experience:

- **Escort:** The attacking team attempts to move the payload from point A to point B, through enemy territory. The defending team attempts to stop the other team from advancing as best as they can until time runs out or the attacking team reaches the destination.

- **Assault:** The attacking team attempts to capture two points across the map by taking control of them as the defending team fights to maintain control. The defending team must protect at least one objective from getting fully captured.

- **Hybrid:** This game mode begins with the attacking team trying to capture a point (assault), then escorting the captured payload to the destination (escort). The rules for each are the same.

- **Control (King of the Hill):** Both teams simultaneously fight over control of objective areas for best 2/3. When one team "controls" the objective, their capture percentage increases slowly towards 100%. Whoever gets it to 100% first, wins. Areas switch within maps after each round.

For each of these game modes, there are multiple maps associated with each. Each map has a unique layout, however, the objective remains the same within a mode. We did not investigate differences across modes, as it is not expected that there would be differences.

4.2.5.3 Why Overwatch?

Overwatch was chosen as the game environment for several reasons. First, the *game complexity mirrors the real world*. Games have time and performance pressures which force players to manage tradeoffs. Each game is different, and players need to be able to both rely on their past experiences and evaluate the situation independently. Teammates are *interdependent* and have to engage in adaptive decision making. The game itself involves two teams who may be ad hoc or pre-made, which is common across many different fields. Regardless of how the teams were formed, they must come together, maintain common ground, and execute in order to accomplish objectives. No one player can produce all the success of a team, and each role on the team depends on the performance and functionality of others. Players engage in *cognitively difficult tasks* as will be enumerated in from the task diagram. *Overwatch* presents players with both time and performance pressures, which makes effective decision making even more critical. Players are also required to manage their resources (i.e. abilities), communicate efficiently with teammates, and collaborate in both planning and execution of activities. *Overwatch* also provides players with an adversarial environment, and challenges players to think, adapt, and act within a quick window. Overwatch challenges players within the game to *perform under time and performance pressures* in dynamically changing scenarios. Players must rely on their gameplay experience to deal with uncertainty and unclear goals, and accomplish their objectives with multiple other players within the constraints and rules of the game. These concepts are not new to cognitively complex work environments (Zsambox & Klein, 2014). The phenomena happening in Overwatch are consistent across other domains, and have been attributed to naturalistic decision making. All of these qualities of *Overwatch* make it a good first demonstration and test environment for The Tracer Method.

4.3 Procedure

Participants were brought into the lab for two sessions, separated by at least one day. During Session 1, participants were consented and briefed on the overall purpose of the study, and each session's respective schedules. In Session 1, participants were interviewed about their specific experiences and decision strategies in *Overwatch* using a developed CDM interview guide (see Appendix A). Participants were asked for additional verbal consent to record the interview. The interview was recorded using OBS. Participants were first asked questions regarding their general *Overwatch* experience, then asked about the most difficult tasks in their role of specialty, and finally were asked about specific games of *Overwatch* from their bank of games.

Session 1 took no more than 1 hour. After Session 1, participants were debriefed and given instructions for Session 2 of the study. As stated earlier, Session 2 occurred at least 1 full day after a participant's Session 1 interview. Participants were allowed to bring their own gaming equipment (i.e. keyboard, mouse, mousepad, headset with mic, controller), or used the game set up in the lab. Participants were permitted to engage in 1-2 warm-up games or practice sessions to prepare them for competitive mode. Participants were encouraged to play with the same friends that they normally do in competitive (if any), however, no data from teammates is reflected in this dataset. The overall study procedure is depicted in Figure 6.

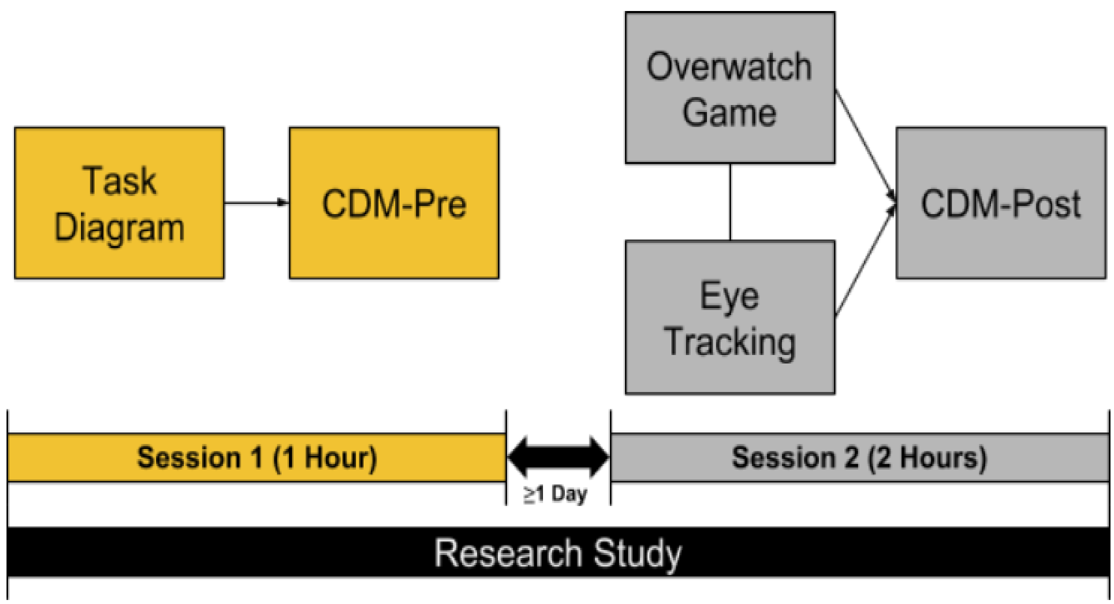

Figure 6 Overall study procedure.

The base interview questions were the same for both the pre and post CDMs (see Appendix A), however, the contextual reference (the game they just played) and the follow up questions that arose are different. These questions are reflected in the transcripts of the interviews and were noted by the researcher. "Incidents" for the Post-

CDM are the decision points determined from the game they had just played. Sweeps 1 and 2 were completed before the participant viewed their game footage. Beginning with sweep 3, the researcher guided the participant through the game footage, focusing on the critical decision points previously established. Providing this visual prompting, we addressed suggestions that Cooke & Cuddihy (2005) and Elling et al. (2011) identified in their papers. Each critical incident was discussed and interviewed in depth and the process repeats for each incident identified. Session 2 took no more than 2 hours total.

4.4 Coding Development

Two independent raters (one author) coded the critical decisions, cues, and courses of action elicited from both interviews. The raters were both experienced Overwatch players with about 850 hours of total gameplay. The general coding procedure involved each rater independently coding the segment of data, meeting together to compare codes, and discussing disagreements to reach full consensus. The coding schemes were developed and finalized over multiple passes through the data.

4.4.1 Incidents

In this study, a game of *Overwatch* is considered an incident. Incidents were valid if the description of the incident included specific events that could fall on a timeline and be recreated temporally. Any incident that did not describe specific elements of an event that could be recreated or visualized on a timeline, was not used in the proceeding coding (Table 1). All post-CDM interviews yielded valid incidents, as they referred to a game that they had just played. All CDM interviews were transcribed and were analyzed from the transcripts, timelines, and interviewer notes.

Table 1 Example incidents from the CDM-Pre.

Usable Incidents	Unusable Incidents
P5: *Playing Reinhardt on Anubis and the team had a dive comp. We quickly captured the 1st point. There was a period of trading kills back and forth. I got my ultimate, communicated my intention to Ult as we went around the left side. I decided to Ult and I got 6 people in it. We got a team kill, captured the 2nd point, and won.*	**P12:** *Diving into the enemy team without a healer (as Dva or Winston) and the bubble doesn't last long enough (or gets destroyed) and I die.*
P7: *We lost the first point (2 CP) and decided to retreat. Their team was getting small picks and getting ult charge. Our team got caught in a Zarya ult and I ulted. Their DPS burned ults, then once theirs ran out, our team used ults and wiped their team.*	**P13:** *I was playing DPS and my team was getting really frustrated because we weren't pushing/capping the point. Throughout the game I was flubbing my ult.*

The unusable incidents do not provide identifiable critical decisions or timelines. For example, P13's incident describes their frustration during a game when their team was not initiating or moving towards the objective. Even with multiple follow-up questions that were reworded, the participant was unable to describe specific critical decisions or a timeline of the game. If the interviewer was unable to elicit any more supplemental information that could be developed into a timeline, the incident was deemed unusable.

4.4.2 Critical Decisions

The main coding scheme for the critical decisions broke down into: Tactical, Operational, and Strategic decisions. This framework was adjusted from the U.S. military theory of war (USAF, 1997). Each level has a different temporal element as well as differing goal scopes. This framework was chosen because it is specifically concerned with the planning and execution of strategy during different levels of conflict. Despite being written for the context of military, it fits well in *Overwatch* as players are required to strategize and execute throughout the game with various time pressures.

- **Tactical decisions:** fast paced, reactive, and specific decisions often relating to specific objective. For example, P3 (Tank)'s second critical decision was to: "*use Earthshatter (ult) during [the enemy] Moira's ultimate."* This decision was in direct response to an enemy's action, and the participant identified it as more reactive.
- **Operational decisions:** higher level and more thought out decisions within a shorter window of time that may also involve delegation and short-term planning. An example from P16 (DPS) was when he described a decision to: *"[Use] High Noon on point 2.*" This participant specifically described referencing their knowledge that this ability could be used to zone enemy players out, and their team was very close to finishing the escort.
- **Strategic decisions:** significantly more planned out and may have been executed over a longer period of time (or several steps). For example, P15 (Support) described their first critical decision as: "*gauge[ing] the flow of the team early; [and use the] time to gather strategies and tendencies of the team.*" P15 admitted that this decision was critical in determining his playstyle for the remainder of the game, and affected the priority of targets.

This coding scheme supports one perspective of our research question investigating differences in decision making from CTA, ET, and Tracer Method data. Another coding scheme was developed using an initial card sort with the existing Macrocognitive framework (Klein et al., 2003), and was reduced to four categories: Sensemaking, Situational Awareness, Coordination, and Managing Uncertainty and Risk.

- **Sensemaking:** generally, how people are making sense of their environment and behaving accordingly; overall, "what do the elements of the environment tell me?" P10 (DPS) discussed his critical decision to "*use [my] ult in a correct position."* This decision required a general understanding of the physical environment and factors from both enemy and ally teams.
- **Situational Awareness** (not related to Endsley's framework)**:** acting on specific elements of the situation or environment that are of higher priority or impactful; this can include detecting problems or managing attention. P2 (Support) decided to, "*ult to save [ally] Mercy who had been ulted by [the enemy] Soldier*." This decision involved P2's ability to detect a problem and make a decision to prevent a bad outcome (ally Mercy dying).
- **Coordination:** planning, communication, or articulation of data or decisions requiring the input or action of others. P11 (Support) made a decision to, "*combo ultimate with Genji."* P11 described this decision as a coordinated decision to initiate advancement onto the second objective.
- **Management of Uncertainty and Risk:** the weighing or structuring the hierarchy of consequences or actions that may or may not involve degrees of

uncertainty. P6 (Tank) articulated their risky decision to, *"[when I lost mech] Tried to get it back quickly to buy time for [my] team to respawn (versus retreating).*" They made the decision to risk dying in order to stall for more time for their back-up to arrive.

This coding scheme also provided a different perspective in addressing our research question examining decision making across multiple methodologies.

4.4.3 Cues

Cues were coded into two main categorizations, both for understanding The Tracer Method output.

4.4.3.1 Information Type

Cues were first categorized by the type of information they represented or elicited. The 5 categories included: Communication, Enemy Status/Action, Game Environment/Objective, Game Sense, and Teammate Status/Action. Table 2 contains category descriptions and several examples for each.

Table 2 Description and examples of cue categorizations.

Category	**Definition**	**Examples**
Communication	Cues associated with verbal or non-verbal communications of information from teammates	*Ana prompted nano; Comms to get out*
Enemy Status/Action	Cues associated with an enemy hero's health, damage, abilities, location, actions, or strategies	*Orisa wasn't in position to get him; Knew enemy Lucio + Dva dead or spawning*
Game Environment/Objective	Cues associated with general game environment entities (i.e. line of sights, specific landmarks on maps) or the objective at hand	*Objective tracker; Death symbols on point*

Game Sense	Cues associated with gamers' "intuition" or that relied on prior game knowledge or strategy (often from expertise)	*Ultimate works as zoning tool; Team at disadvantage (positional)*
Teammate Status/Action	Cues associated with an ally (or self) hero's health, damage, abilities, location, actions, or strategies	*Ana (team) location; Genji's dash in air (team)*

4.4.3.2 Mappability and Confirmation

Cues provided by the participant from their CTA interview were determined to be mappable or not mappable regarding eye tracking. Mappable cues were defined as: cues that could be identified to an eye gaze pattern within the visual game environment. Mappable cues have a source for the information (specified or not by the participant). For example, Mercy's health can be elicited from several visual cues (i.e. her health bar, the critical health icon, callout in chat). Unless the participant specified eliciting Mercy's health through an audio cue (such as her "I'm under attack!" voiceline), it would be considered mappable due to the visual source. Audio cues were automatically identified as not mappable. Cues that relied on prior game knowledge, or "game sense" were not considered mappable. For example, Figure 7 is a screenshot from P11's game, with the blue circles representing fixations. From their interview, P11 identified six cues: Reinhardt had low health, Self-health, Team cues, Junkrat having to cover distance, Junrat's grenades, and Reinhardt's health. Reinhardt's low health, Reinhardt's health, Self health, and Junkrat's grenades are mappable because there are defined sources in the environment from which the information can be elicited. Junkrat having to cover distance and team cues are not considered mappable because P11 identified team cues as being from audio comms, and Junkrat having to cover distance constitutes a game knowledge cue, even though you could identify his distance in the environment, knowing the hero's mobility constitutes game sense. Figure 7 is a screenshot with potential sources outlined in green and orange.

Figure 7 Screenshot from P11's game in which they described six cues, including: Reinhardt's health and self-health. Both of these are examples of mappable cues, with the potential sources for Reinhardt's health outlined in green (middle and middle-left), and self-health (bottom left) outlined in orange.

Mappable cues were then coded for confirmed and not confirmed status based on the ET. Confirmed cues were defined as: having fixated on an element which would provide them that information or fixating on an element itself indicated by a cue (i.e. Mercy's health from the critical indicator, or Mercy's health bar itself) within the 20 second critical decision interval. To constitute confirmed cue, the participant must fixate on it or within reasonable proximity to be considered within the foveal field. Thus, if a participant's gaze trail passed through the cue element but does not fixate, it is not considered confirmed. Any cues that were outside of the 20 second critical decision window were also not confirmed (and were not considered in the analysis). For example, in Figure 8, using a different scenario, P10 articulated using eleven cues. The mappable cues included: self-health, enemy location, did enemies see him, opponents pushing onto point, Rein no shield, and what the enemy was doing. In this screenshot (Figure 8), self-health would be confirmed, as the participant fixated on the cue source within the decision window.

Figure 8 Screenshot from P10's game in which they described self-health as a cue. Self-health is both mappable and confirmed by ET, outlined in green.

4.4.4 Courses of Action (CoA)

Courses of action (CoA) were categorized by the type of "action" they represented. The 3 categories included: Action, Planning, and Inaction. Table 3 contains descriptions and examples of each category.

Table 3 Description and examples of CoA categorizations.

Category	**Definition**	**Examples**
Action	Courses of action associated with taking another "action" or intensifying the current action	*Charge Moira; Gone into room after Symmetra*
Planning	Courses of action associated with strategic, generalized, or coordinated efforts	*Gone a different route; Regroup*
Inaction	Courses of action associated with inaction, inhibition, lack of change, or retracting of a decided action	*Let Reinhardt do it; Stayed on Zarya*

5 Chapter 5: Results

This chapter focuses on the findings directly related to MR 1 and MR 2 research questions. The chapter opens with a general description of the data analysis procedure. To evaluate The Tracer Method, we begin by analyzing the data from the CTA and establishing unique contributions of the CDM. This includes analyses from both the CDM-Pre interview, the Task Diagram, and the CDM-Post interview. The CDM-Pre are interviews pulling from participants' extensive Overwatch game play experience. In contrast, the CDM-Post is focusing only on the specific team game that they played in the lab, a game that also involved eye-tracking. This distinction between the CDM-Pre and CDM-Post allows us to explore if there are differences in the types of incidents, decision or cues. After the CDM descriptives are outlined, the AOI and PVAL status analyses of the ET data is described. The general ET fixation and visit counts, fixation durations, and order will be described by decision type and role, however, there were no specific hypotheses related to this measure. The Tracer Method section will describe the findings related to both MR 1 and MR 2 separately.

5.1 Data Analysis Plan

The general data analysis procedure occurred over several steps. CTA data was analyzed first, with over 30 hours of interview audio to sift through. The relevant portions of the audio recording of the interviews were transcribed and additional notes were taken from the audio. Two independent raters (one author and one external to the study) coded the critical decisions, cues, and courses of action over the course of multiple sessions.

To prepare for the ET analysis, for each critical decision, one of the authors had to segment off a 20 second window (10 seconds before and 10 seconds after the decision action). This time frame was chosen because it captured the elements of each decision and prevented an overlap in decision windows. The AOIs had to be overlaid on each participants' game footage and AOI groups were created to increase the speed of ET data processing. ET data incorporated: the number of fixations, average fixation duration, and number of visits to AOIs, which were calculated by the Tobii Pro Studio software. PVAL coding required the researcher to subjectively code 1,766 fixations by what was being fixated on. This was an extensive process, which took several weeks. This analysis is not necessary for every environment, or even game. We chose to do this analysis because it was important to further unpack the large number of fixations in the PVAL.

The new data from The Tracer Method included: a between-ness visualization and the PVAL Analysis. To determine if a cue was confirmed, one of the authors re-watched the 20 second critical decision windows of game footage and analyzed whether the cues mentioned in the CTA were fixated on or not. The performance metrics from the game were recorded across participants, but the data was not analyzed as it is irrelevant to the study focus. The data from these multiple methodologies made the analysis process longer and more complex, however, it was focused only on the most important parts of the data (between 40-80 seconds of a 30 minute game). Multiple sweeps were required to ensure accuracy and consistency. The process spanned over multiple months.

5.2 Cognitive Complexity of *Overwatch* (Task Diagram)

The Task Diagram provided a general understanding of what was happening in Overwatch and what tasks were cognitively complex. The tasks were put into themes, and across all roles, there were both consistencies and differences (Figure 9). The output from the Task Diagram is helpful to gain an understanding of the domain, in this case, Overwatch. It also provides and overview of the kinds of cognitive activity that is both overlapping and unique to each role. Across DPS, Tank, and Support players, they consistently identified difficult tasks that involve: resource management, tracking/monitoring of players, and strategic positioning.

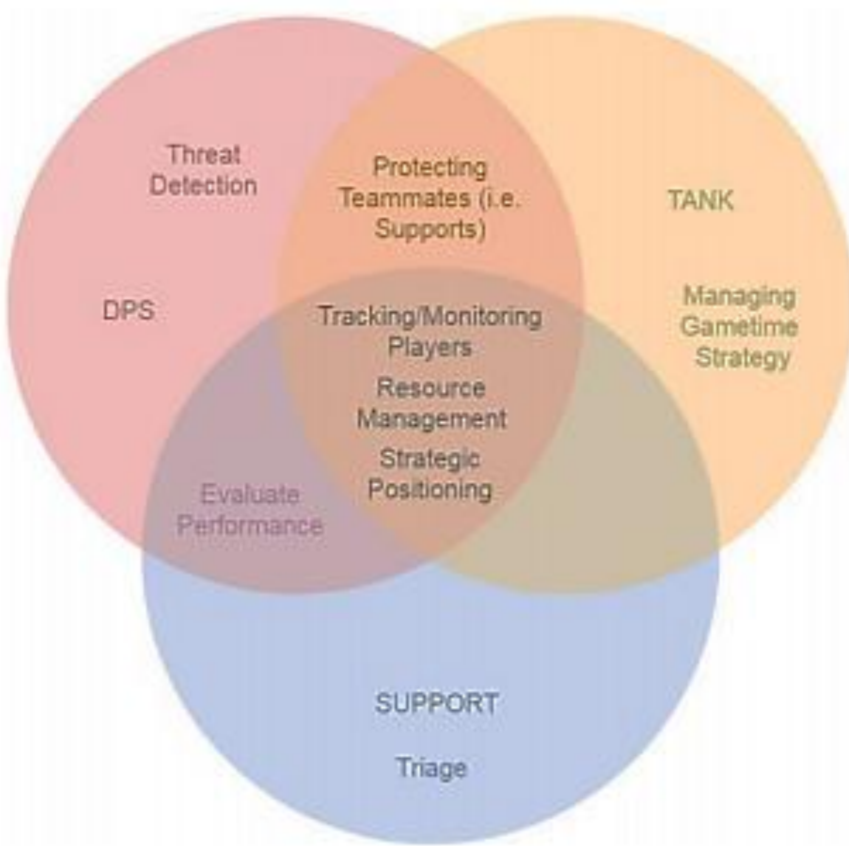

Figure 9 Venn diagram of the cognitively difficult tasks across roles.

Table 4 provides specific examples from the Task Diagram for commonly complex tasks. What is clear across these examples is that each task theme demonstrates how players must manage trade-offs between various game elements (e.g. managing abilities vs. ultimates).

Table 4 Examples of cognitively complex tasks that are common across all roles.

Common Task Theme	**Definition of Category**	**Example from Task Diagram**
Resource Management	How to use what you have (abilities + ult) and when to use it (abilities + ult)	*"Do you want to use [the ability or ultimate] for kills or zoning?"* **P12, Tank**
Tracking/ Monitoring Players	Monitoring location (or health) of players (both ally and enemy)	*"Some heroes that you need to monitor are off-screen."* **P14, Support**
Strategic Positioning	Positioning of self, relative to others	*"Prevent enemy flankers and disrupt their LOS."* **P17, DPS**

DPS players typically described their most difficult tasks as involving threat detection. These tasks require knowledge of both enemy and ally team compositions, determination of threat priority, and primary execution of the threat. Tanks, on the other hand, identified the most difficult tasks related to managing game-time strategy. These tasks require adaptive decision making regarding their playstyle or maneuvers (offensive and defensive) based on both teammate and enemy actions or tendencies. Unique to Support players, determining who the priority is to receive healing (Triage) is both difficult and critical. This requires understanding of team sustainability and importance of the status of each member. Table 5 describes the cognitively complex task themes (with examples) that were different (or exclusive) across roles.

Table 5 Examples of cognitively complex tasks that are different across roles.

Role	Task Theme	Definition	Example
DPS	*Threat Detection*	Identifying and deciding on focus targets (i.e. who is the biggest threat)	*"Decide which [target] is best and killing them fast or maim them."* **P9, DPS**
Tank	*Managing Gametime Strategy*	Adapting playstyle and making decisions associated with general game-time events	*"Have lots of health to manage; difficult to track."* **P8, Tank**
Support	*Triage*	Priority (hierarchy) of healing targets	*"How much healing or sustain does the team have?"* **P7, Support**

The Task Diagram was a first look at the events and tasks involved during Overwatch games. This data was helpful in understanding the level of complexity of the game, and what type of data we may elicit from the proceeding CDM interviews.

5.3 Critical Decision Method

5.3.1 Critical Decision Method Pre (CDM-Pre)

The CDM Pre interviews were analyzed to identify incidents and their corresponding elements: critical decisions, cues, and courses of action. The 17 interviews produced an overall 27 critical decisions across 25 incidents total. There is an average of 1.1 (SD=0.94) decisions per incident. Of the 25 incidents, 17 incidents (one per participant) were chosen for the CDM sweeps 2-4. Across all roles, participants identified a total of 106 cues and 21 CoAs across 16 critical decisions. On average, participants identified 6.6 (SD=4.06) cues and 1.2 (SD=0.83) CoAs per decision. Table 6 describes the breakdown of these 17 incidents, critical decisions, cues, and courses of actions across roles.

Table 6 Number of incidents, critical decisions, cues, and courses of action by role from the CDM Pre.

Role	# Incidents	# CDs	# Cues	# CoAs
DPS (N=6)	6	4	39	7
Tank (N=6)	6	7	39	9
Support (N=5)	5	5	28	5

5.3.2 Critical Decision Method Post (CDM-Post)

The CDM-post is the key connection to the eye tracking analysis, however, this section will just describe the unique contributions of the CDM-Post interview. The Session 2 interviews yielded 39 critical decisions across 14 incidents, at a rate of 2.8 decisions per incident. The 39 critical decisions distributed into 10 Strategic, 14 Operational, and 15 Tactical decisions. For the macrocognitive themes, the critical decisions were placed into 13 Sensemaking, 13 Situational Awareness, 6 Coordination, and 7 Managing Uncertainty and Risk decisions. In addition, the CDM interviews elicited 283 total cues, with an average of 7.3 (SD=3.12) cues per decision. Of the 283 cues, 225 cues were visual cues (79.5%) and 58 were audio cues (20.5%). Participants identified 49 total courses of action across the 39 decisions, with an average of 1.3 (SD=0.85) CoAs per decision. Table 7 describes the distribution of critical decisions, cues (and type), and courses of action across Strategic, Operational, and Tactical decision types.

Table 7 Distribution of critical decisions, cues, and courses of action across decision types for the CDM Post.

Decision Type	# CDs	Total Cues	# Visual Cues	# Audio Cues	CoAs
Strategic	10	82	66	16	14
Operational	14	86	71	15	17
Tactical	15	115	88	27	18

We also investigated the distribution in cue categorization across decision types. There were statistically significant main effects of decision type on number of communication cues $F(2,36)=3.764$, $p<0.05$ and game sense cues $F(2,36)=6.134$, $p<0.01$. The pairwise comparisons revealed the driving differences between Strategic-Operational $t(14)=2.6$, $p=0.02$, and Tactical-Strategic decisions $t(13)=-2.57$, $p=0.02$. Strategic decisions had more game sense cues than Tactical and Operational decisions. A Pearson's Chi-squared test was significant on the distribution of cue category across decision type $\chi2$ (8) =30.57, $p<0.001$. The means and standard deviations can be found in Table 8, and the respective percentages in Table 9.

Table 8 Means and Standard Deviations for cue categorizations across decision types.

	Strategic	**Operational**	**Tactical**
Communication	1.2 (2.0)	0.21 (0.43)	0.33 (0.62)
Enemy Status/Action	1.4 (1.2)	2.1 (1.2)	2.2 (1.5)
Game Environment/Objective	0.5 (0.53)	0.57 (1.1)	1.5 (1.7)
Game Sense	3.5 (2.1)	1.6 (1.3)	1.7 (1.1)
Teammate Status/Action	1.6 (1.4)	1.7 (1.8)	2.0 (2.1)

Table 9 Percentage distributions for cue categorizations across decision types.

	Strategic	**Operational**	**Tactical**
Communication	14.6%	3.5%	4.3%
Enemy Status/Action	17.1%	33.7%	28.7%
Game Environment/Objective	6.1%	9.3%	19.1%
Game Sense	42.7%	25.6%	21.7%
Teammate Status/Action	19.5%	27.9%	26.1%

5.4 Eye Tracking

5.4.1 Area of Interest Analysis

Eye tracking was analyzed using fixation counts, visit counts, and fixation durations across the 10 AOIs. Zeroes were counted in fixation and visit counts, however, fixation durations associated with a zero count was given an *NA* value and were removed from the analyses. Fixation counts included all individual fixations within pre-determined AOIs. Visits are defined as all fixations within one unique visit to an AOI. For example, if a person fixates 10 times within one AOI, that would constitute one visit if they did not fixate outside the AOI during those 10 fixations. Because previous literature suggests that the majority of fixations occur in the PVAL (reticle), calculating visits would maintain the distribution pattern, but would reduce the noise in the data. Thus, the proceeding analyses calculate both fixation and visit counts.

A one-way ANOVA indicated statistically significant main effect of AOI on fixation counts $F(9,170)=282.09$, $p<0.0001$. There was also a statistically significant main effect of AOI on average fixation duration $F(9,46)=2.433$, $p<0.050$. The statistically significant differences were between the PVAL-Ability CD, PVAL-Death Tracker, PVAL-Health/On Fire, PVAL-PVAL Left, and PVAL-PVAL Right. In all of these pairwise comparisons, except for the PVAL-PVAL (left and right), the average fixation duration was longer in the PVAL. Figures 10 and 11 depict the distribution of average visit counts and fixation durations across each AOI respectively (averaged across all critical decisions). This finding is consistent with ET literature that concludes that longer fixation durations are an indicator of required processing time for more complicated stimuli (Poole & Ball, 2006).

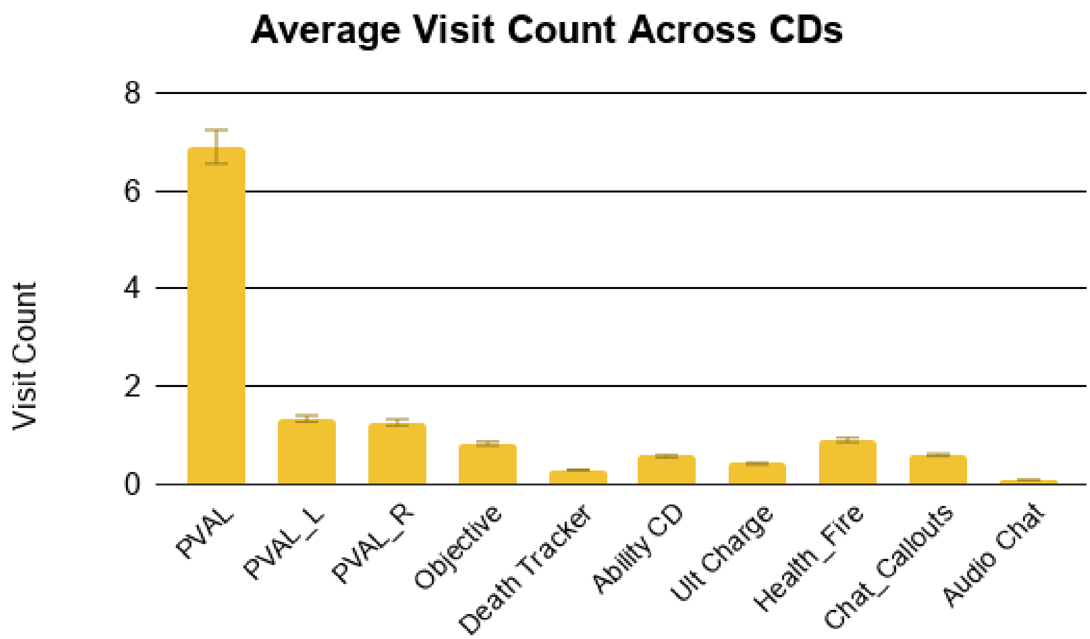

Figure 10 Average visit count across AOIs.

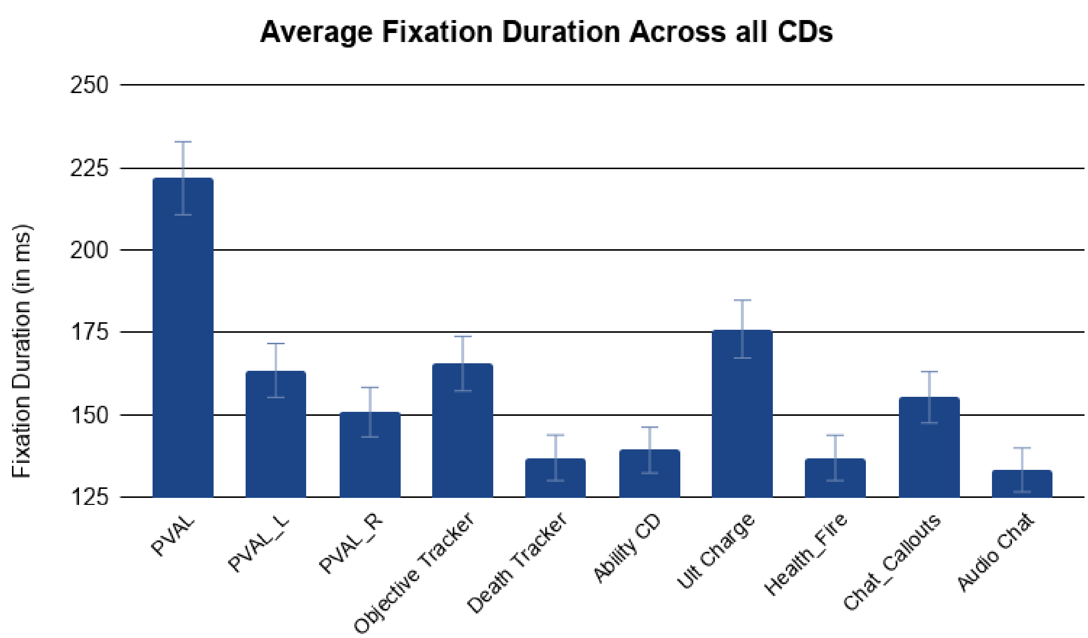

Figure 11 Average fixation duration across AOIs.

5.4.2 PVAL Status Analysis

Fixations were coded into PVAL and non-PVAL AOIs (including PVAL_L and PVAL_R as these were indicators of scanning behaviors, but were not part of the PVAL). The benefit of combining all of the non-PVAL AOIs is that it is still an indicator of fixation or visit diversity, without increasing the number of levels. There were statistically significant main effects of PVAL status (AOI in PVAL or not) on fixation counts $F(1,388)=2731$, $p<0.001$, fixation durations $F(1,179)=27.46$, $p<0.001$, and visit counts $F(1,385)=746.35$, $p<0.001$. Thus, the location of the AOI (inside or outside of PVAL) affects fixation counts, durations, and visit counts. To ensure simplicity and cleanliness in data analysis, PVAL status (PVAL vs. Not PVAL) is used for all proceeding analyses in Chapter 6.

5.4.3 AOI Order Analysis

Eye Tracking also provides the opportunity to elicit frequent orders of visits. Visits were calculated by collapsing across fixation counts. We decided to analyze the most common 2-AOI visit pattern orders across decision types and roles. The most common patterns for each decision type and role are described below.

Decision Types:

- *Strategic Decisions (33 unique transitions)*
    - **Top:** Death Tracker→ Ult Charge (n=6)

- *Operational Decisions (38 unique transitions)*
    - **Top:** Ability Cooldowns/Ammo→ Objective Tracker (n=5)

- *Tactical Decisions (42 unique transitions)*
    - **Top:** Ability Cooldowns/Ammo→ Death Tracker (n=8)

Roles:

- *DPS (34 unique transitions)*
    - **Top:** Ability Cooldowns/Ammo→ Death Tracker (N=6); PVAL/Reticle→ Ult Charge (N=6), and Death Tracker→ Ult Charge (N=6)

- *Tank (37 unique transitions)*
    - **Top:** Death Tracker→ Ult Charge (N=6)

- *Support (33 unique transitions)*
    - **Top:** Objective Tracker→ Ability Cooldowns/Ammo (N=8)

Transition frequency ranged from 1-8 instances across all decision types and all roles. Tactical Decisions having the highest frequency and the most unique transitions. Supports had the highest frequency but Tanks has the most unique transitions. We did not specifically hypothesize a difference across decision types or roles, and within the context of eye tracking, we cannot speculate much more about the meaning of these transitions. The Tracer Method provides a way to make justifiable speculations about the meaning of this data, as evidenced in the between-ness visualization described in the next section.

5.5 The Tracer Method

This section will outline the specific findings from The Tracer Method. First, as consistent with past research, we will discuss overlapping and different (or unique) information. Moving in a new direction, we will discuss the new information that emerges from The Tracer Method.

5.5.1 Overlapping and Unique Information

Past literature has suggested that the overlap between verbal protocol and process tracing methods (i.e. concurrent verbal protocol) and eye tracking is not perfect, but is at least 75% (Cooke, 2010) and maxes out at 98% (Rhenius & Deffner, 1990). Based on their chosen methodology, it can be assumed that the majority, if not all, of the information from the interview can be directly mapped to ET. In our study, the maximum percentage of overlap is 60% (170 mappable cues/283 total reported cues from the CDM-Post), which is significantly lower than what has been found in the past, which was 77% of observable verbalizations from Cooke's (2010) study. This maximum assumes that all 170 mappable cues were observable in the eye tracking, which was obviously not the case. The equivalent calculation of the amount of overlapping information is the ratio in this study is the number of ET confirmed cues to the number of total mappable cues. Of the total 170 mappable cues (of 283 total from the CDM-Post), 59% (100 cues) of the cues were validated by the ET within the decision window. In Elling et al, (2012) their percentage of "different" information (27%) was calculated through the silences during the concurrent think aloud. In this study, the percentage of unique information is much higher (41%).

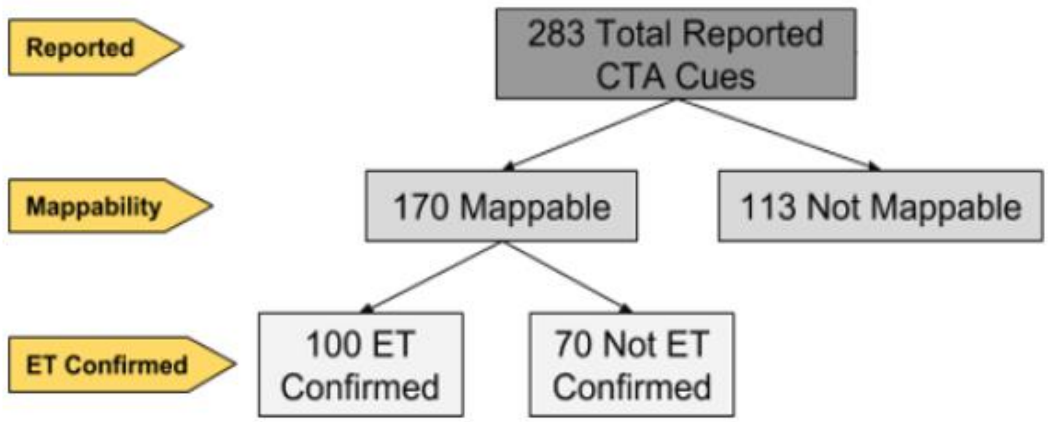

Figure 12 Reported CDM cues distributed across mappability and ET confirmation.

The large reduction in overlap could be due to a number of factors: 1. The methodologies for eliciting information (concurrent verbal protocol vs. CDM) are very distinct, 2. The environments (website vs. FPS video game) that the methods are being utilized in affect the type of information being elicited, or 3. The focus of the research questions affects the unitization of the data. It is likely that all 3 of these reasons are at play, however, it is important to note that a systematic method like CDM has not been combined with ET in a dynamic and complex environment like a video game, thus, we can only compare the amount and type of information elicited from the combination of CTA and ET from what has been done in the past. These findings suggest that there is a great benefit in the combination, and that there is power in the new information, which is described in the next section.

5.5.2 New Information

To evaluate the extent to which these methods provide new information when used concurrently, we can define the information that can only be elicited with the help of both. On the surface, both of these methods produce completely different information:

CTA produces cognitive and ET produces behavioral data. We argue that eye tracking requires a supplemental method to provide meaning to the data. Without a guided question or context for investigation, the ET analysis can only provide various eye movement data points, which have little to no meaning. However, this does not mean that eye tracking is inferior, but rather, it is critical to developing a systematic context for the ET in order to elicit the most meaningful data.

There is data that can potentially be elicited using a bottom up approach (from ET) using the context of the CTA interview (Bednarik & Tukiainen, 2006). Previous research from Cooke and Elling (2011), "silences" from the verbal protocols were analyzed using the same approach. The distribution of mappable and not mappable cues across decision types can be found in Table 10.

Table 10 Distribution of total mappable and not mappable cues across decision types.

	Strategic	**Operational**	**Tactical**
Reported Mappable	40	52	78
Not Mappable	42	34	37

A Pearson's Chi-squared test was significant on the distribution of mappable and not mappable cues across decision type $\chi 2$ (2) =7.25, $p<0.05$. The effect of decision type on mappable and not mappable cues further adds support for the new information hypothesis. There would be no way to determine a difference in mappable and not mappable without the context or information from the CTA. On the other side, there is benefit in having the potential for cross-validation from ET for the mappable cues.

5.5.2.1 Between-ness Analysis

Following up the order of fixation analysis, we conducted a between-ness analysis. This analysis describes the relationship between eye movements in a network. It determines the level of centrality between nodes (AOIs). In other words, it describes which elements of the UI are more central (or important) in visit transitions throughout decisions. Figure 13 (next page) provides the *Overwatch* interface with each AOI blocked out and labeled. The square representations below map to each of the three roles and the three decision types. Each colored dot represents a node (AOI). The black lines between nodes represent transitions between the AOIs, and their thickness symbolizes the centrality or the strength of the transition. The number of arrows leading to a node represent the relative frequency of transitions, creating a network diagram (bottom). Figure 13 demonstrates the network diagram for both decision types and role.

Beginning with decision types, Tactical decisions have the strongest link between PVAL/Reticle and Ability Cooldowns/Ammo. This can seem somewhat expected on the surface, as both Tactical decisions and Ability cooldowns by nature are strictly time central. More deeply, making decisions that have shorter time windows would require a rapid understanding of the current resources or state of the environment. This explanation would also carry over into unpacking the strong traditional relationship between the PVAL/Reticle and Ult charge. Ultimate abilities are generally defined as strong skills or player controlled “events” that tend to be impactful in game. In other words, Ults can be “game changers” or can strongly impact the outcome of a fight or game. This resource is not as readily available as basic abilities or ammo, however, knowing the status of it with a time critical decision could strongly impact the outcome. Interestingly, this connection between PVAL/Reticle and Ult charge isn't just strong for Tactical decisions, but carries over into Strategic decisions, and less so with Operational decisions. Strategic decisions have a much longer time window and larger overarching goal orientation, thus, being able to predict the timing of Ultimate availability would be important to being able to plan accurately. The specific transitions varied slightly by decision type, however, Operational decisions overall showed the weakest transitions between AOIs compared to Strategic and Tactical decisions.

Specifically for roles, both DPS and Tank show greater transitions from PVAL/Reticle to Ult charge. Both of these roles generally have offensively oriented Ultimates (some can be used defensively), and also have offensive responsibilities (DPS for getting eliminations and Tanks for initiating team fights). These role features are evident in both their Task Diagram and looking at the transitional probabilities. Although there was no difference in distribution of decision types across roles, the general strength pattern is similar between DPS and Tactical. Supports differed slightly, demonstrating greater transitional probability for PVAL/Reticle to Ability Cooldowns/ammo and Death Tracker to Ability Cooldowns than the other two roles. As Supports described (unanimously) their most difficult task during the game to include Triage, their abilities are typically associated (and necessary) with completing the tasks required of their role. An example that can be tie this analysis to the CDM-post data is: Tanks elicited the most communication and team status cues among the roles, and the between-ness visualization demonstrates stronger links between the PVAL and the Chat/callouts and the PVAL and the audio chat than other roles. It is difficult to speculate much more on the interpretation of this data, however, the CDM can provide some insight into potential ties to the cognitive processes.

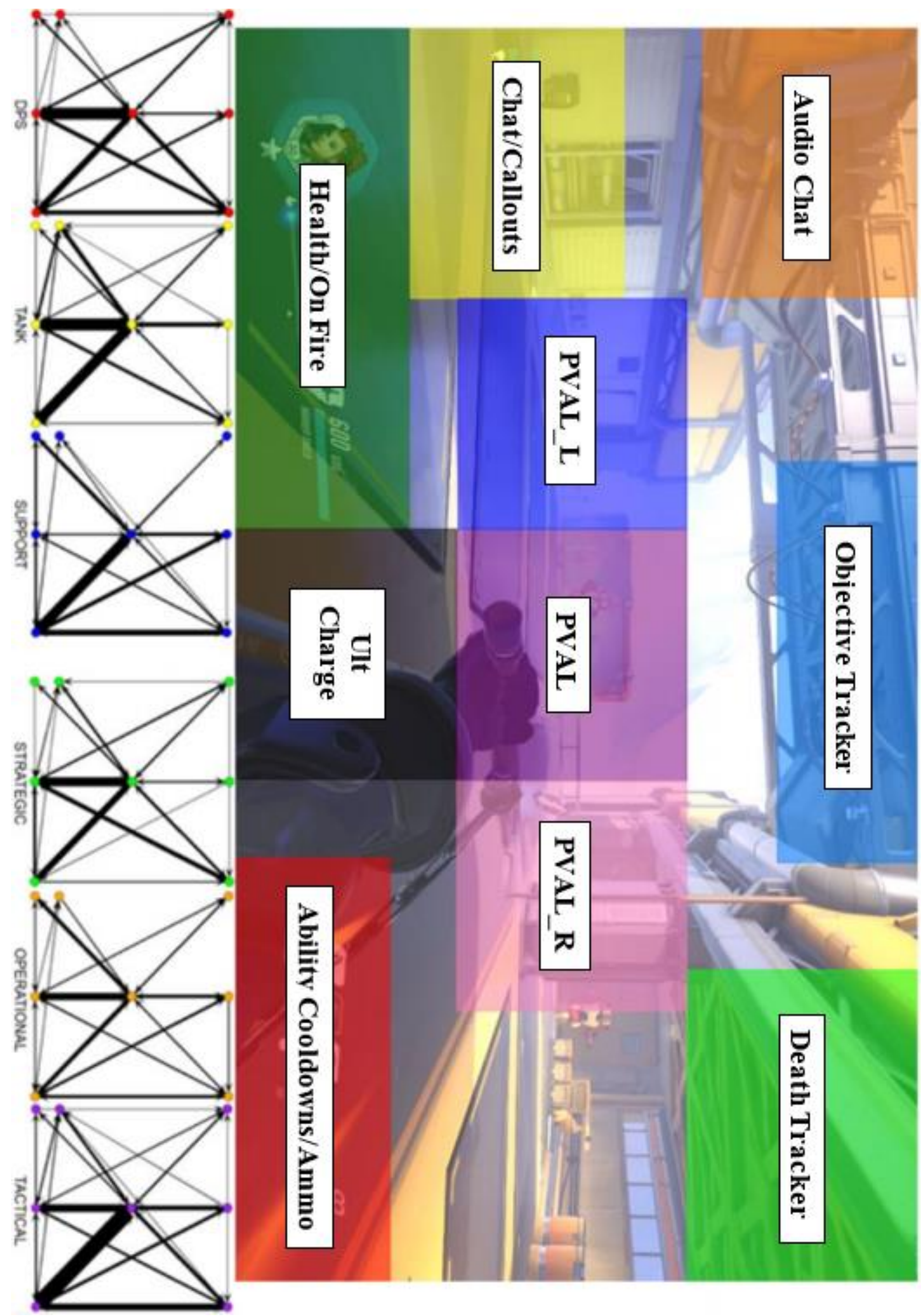

Figure 13 Visualization of Between-ness

5.5.2.2 PVAL Analysis

Following up the between-ness visualization analysis, The Tracer Method is able to demonstrate the cognitive complexity of a FPS through the PVAL Analysis. As previous research in FPS has suggested, the majority of eye movement activity occurs in the PVAL/Reticle. Researchers may be discouraged to use or study FPS games because of this, however, The Tracer Method was able to tease apart the 1,766 fixations in the reticle into what information is being elicited. The coding scheme was taken from the cues categorization, and slightly adjusted to fit the reticle (and to include oneself). From the CDM, we know that people can elicit the same "type" of information from a variety of sources, some of which are in the PVAL and others that are not. From previously discussed findings, we know that the decision type affects individuals' behavioral eye patterns between PVAL and non-PVAL AOIs. This analyses demonstrates that across decision types, the distribution across type of information being focused on in the PVAL is different $\chi^2(6, N=1,744)=89.4$, $p<0.0001$. The Chi Square for fixation category is also significantly distributed across decision types $\chi^2(6, N=1,744)=89.4$, $p<0.000001$. Twenty two fixations could not be coded due to various errors. The distribution of fixation category across decision types can be found in Table 11.

Table 11 Frequency of PVAL fixation information type across decision types.

	Strategic	**Operational**	**Tactical**
Self	8 (0.05%)	34 (1.9%)	56 (3.21%)
Teammate	163 (9.35%)	247 (14.2%)	145 (8.31%)
Enemy	108 (6.19%)	291 (16.7%)	302 (17.3%)
Game Env./Objective	119 (6.82%)	148 (8.49%)	123 (7.05%)

6 Chapter 6: Other Findings

6.1 Overwatch Role Analysis

We evaluated *OW 1* in both the CDM Pre and CDM Post data. The hypothesis that Tank players will identify the most cues and courses of action was not supported by the CDM-Pre interview data. There was no main effect of role on number of cues $F(2,16)=0.231$, $p=0.796$ or courses of action, $F(2,16)=0.368$, $p=0.70$. This is not surprising, as players may not have recalled important details from the various games. Because they were pulling for their bank of Overwatch games, the data demonstrates more general trends. Table 12 lists the means and standard deviations for cues and CoAs for the CDM Post across role. *OW 1* was not supported for number of cues as there was no difference in number of cues across role $F(2,36)=0.18$, $p=0.84$. However, consistent with *OW 1*, there was a statistically significant main effect of role on number of courses of action $F(2,36)=3.89$, $p=0.03$. The pairwise comparisons indicated that the difference was driven by Tanks having more CoAs than DPS players $t(17)=2.36$, $p=0.03$ and marginally more CoAs than Supports $t(18)=2.1$, $p=0.05$.These findings partially support the hypothesis that Tanks will elicit the most courses of action, but not the most cues.

Table 12 Means and standard deviations for number of cues and courses of action across roles from the CDM Post.

Role	# CDs	# Cues	# CoAs
DPS (N=5)	14	7.4 (2.9)	1.0 (0.68)
Tank (N=4)	11	7.5 (3.9)	1.8 (0.98)
Support (N=5)	14	6.9 (2.8)	1.1 (0.73)

The proceeding research questions were only evaluated with the CDM Post data. To evaluate the *OW 2* question, a Pearson's Chi-squared test was conducted to analyze the distribution of decision type across roles (Table 13). The Chi square was not significant $\chi^2(4)= 3.016$, $p=0.56$, indicating that the distribution of decisions provided in the interviews did not differ across roles. Thus, the Chi-squared test does not support the hypothesis that Tanks will make the most Tactical and Operational decisions and Supports will make the most Strategic decisions, or that the distribution is even different across roles.

Table 13 Distribution of critical decision types across roles.

Role	# CDs	# Strategic	# Operational	# Tactical
DPS (N=5)	14	4	5	5
Tank (N=4)	11	3	2	6
Support (N=5)	14	3	7	4

In evaluating *OW 3* question asking: to what extent do the roles differ in distribution of fixation and visit counts and fixation durations across AOIs, multiple 2 PVAL status x 3 Role mixed factorial ANOVAs were conducted on fixation counts, fixation durations, and visit counts. Tables 14, 15, and 16 describe the means and standard deviations of fixation counts, fixation durations, and visit counts respectively. There was no main effect of role on average fixation counts across AOIs $F(2,384)=0.39$ $p=0.67$. However, the interaction between PVAL Status and role on fixation counts was significant $F(2,384)=5.4$, $p=0.005$. This describes the impact of role and PVAL status on average fixation counts across AOIs. We then evaluated the difference across roles on fixation durations. There was no main effect of role on average fixation duration across

AOIs $F(2,175)=1.6$, $p=0.21$. The interaction of role and PVAL status was not significant either $F(2,175)=0.14$, $p=0.87$. The lack of interaction between these variables indicates a lack of variability in the distribution of fixation durations across AOIs. Finally, we investigated the difference in visit counts across roles to address the uneven distribution across AOIs being driven by the PVAL. There was no main effect of role on average visit counts across AOIs $F(2,384)=1.6$, $p=0.20$. The interaction of role and PVAL status was not significant either $F(2,384)=1.8$, $p=0.16$. The lack of interaction between these variables indicates a lack of variability in the distribution of visit counts across AOIs.

Table 14 Average fixation counts and standard deviations across role.

PVAL Status	**DPS (N=14)**	**Tank (N=11)**	**Support (N=14)**
PVAL	48.5 (SD=11.9)	43.0 (SD=13.0)	43.4 (SD=17.2)
Not PVAL	5.7 (SD=14.9)	5.5 (SD=13.2)	5.6 (SD=13.9)

Table 15 Average fixation durations and standard deviations across role.

PVAL Status	**DPS (N=14)**	**Tank (N=11)**	**Support (N=14)**
PVAL	222.7 (SD=67.9)	201.0 (SD=69.2)	235.7 (SD=102.9)
Not PVAL	161.2 (SD=70.2)	171.1 (SD=86.2)	168.2 (SD=82.3)

Table 16 Average visit counts and standard deviations across role.

PVAL Status	**DPS (N=14)**	**Tank (N=11)**	**Support (N=14)**
PVAL	6.3 (SD=2.5)	7.4 (SD=2.5)	7.1 (SD=3.2)

Not PVAL	1.3 (SD=2.3)	1.4 (SD=2.4)	1.4 (SD=2.4)

To further understand the cognitive engagement in decision making in Overwatch, we also evaluated the difference across roles on distribution of cue (Tables 17 and 18) and CoA categorizations (Table 19). There was no main effect of role on any of the cue types. There was, however, an interaction between role and decision type on number of communication cues $F(4,30)=6.103$, $p=0.001$. The significant interaction may be due to low or zero counts in various cells across decision and cue types.

Table 17 Percentage distribution of cues across type and role

	DPS	**Tank**	**Support**
Communication	7.7%	12.0%	4.3%
Enemy Status/Action	28.8%	30.1%	28.7%
Game Environment/Objective	13.5%	10.8%	19.1%
Game Sense	31.7%	19.3%	21.7%
Teammate Status/Action	18.3%	27.7%	26.1%

Table 18 Means and standard deviations of cues across type and role

	DPS	**Tank**	**Support**
Communication	0.57 (0.65)	0.91 (2.0)	0.14 (0.36)
Enemy Status/Action	2.1 (1.1)	2.3 (1.7)	1.5 (0.1)
Game Environment/Objective	1.0 (1.6)	0.82 (1.2)	0.86 (1.2)
Game Sense	2.4 (1.6)	1.5 (0.93)	2.4 (2.0)
Teammate Status/Action	1.4 (1.3)	2.1 (2.3)	2.0 (1.9)

Table 19 Means and standard deviations for CoA type across roles.

	DPS	**Tank**	**Support**
Action	0.83 (0.58)	1.3 (1.0)	0.42 (0.52)
Inaction	0.17 (0.39)	0.44 (0.53)	0.25 (0.41)
Planning	0.17 (0.39)	0.44 (0.73)	0.58 (0.93)

For the CoA analysis, there was a statistically significant main effect of role $F(2,24)=5.193$, $p<0.05$ on number of Action CoAs. Tanks had more Action CoAs than supports, according to the pairwise comparisons $t(15)=2.7$, $p=0.02$. It is not surprising based on a Tank player's role in game that they would have more Action CoAs than Supports, as they have both offensive and defensive responsibilities. No statistically significant effects were found for inaction CoAs. There was a marginally significant interaction between role and decision type $F(4,24)=2.315$, $p=0.0865$ on number of planning CoAs. The interaction of role and both decision type indicates that there are both game and cognitive forces at play during decision making processes, specifically regarding planning.

6.2 Macrocognitive Decision Codes

To evaluate the *DM 2 and 4* questions, multiple analyses of variance were conducted to analyze the effects of decision type (Macrocognitive) on cues (number and type) and courses of action. For Macrocognitive decision coding, means (with standard deviations in parentheses) are reported in Tables 20, 21, and 22. There was a statistically significant main effect of macrocognitive function on the number of CoAs $F(3,35)=2.99$,

p=0.05. Pairwise comparisons indicated that the difference was driven by more Sensemaking decisions than SA decisions t(22)=3.12, p=0.004. There was no significant main effect of macrocognitive function on the number of cues $F(3,35)=1.65$, p=0.2. Sensemaking decisions had more courses of action than SA decisions, which partially supported the hypothesis that Sensemaking decisions elicited more CoAs than SA decisions. DM 2 hypothesis was not supported for number of cues. There was a marginally significant main effect of macrocognitive function F(3,35)=2.448, p=0.08 on number of communication cues. No other cue types were significantly affected by macrocognitive function. There were no statistically significant main effects of macrocognitive type and Action CoAs F(3,29)=1.2, p=0.32, Inaction CoAs F(3,29)=0.55, p=0.65, or Planning CoAs F(3,29)=0.56, p=0.86.

Table 20 Means and standard deviations for number of critical decisions, cues, and courses of action across Macrocognitive decision types.

Macrocognitive Type	**# CDs**	**Total Cues**	**# Visual Cues**	**# Audio Cues**	**CoAs**
Coordination	6	5.3 (1.4)	4.8 (1.2)	0.5 (0.84)	1.3 (0.52)
Sensemaking	13	8.5 (2.8)	6.8 (2.4)	1.8 (1.1)	1.7 (0.82)
SA	13	7.2 (3.3)	5.6 (2.7)	1.5 (1.0)	0.77 (0.65)
Managing Uncertainty and Risk	7	6.7 (3.6)	5.0 (2.5)	1.7 (1.8)	1.3 (1.1)

Table 21 Means and standard deviations for cue types across Macrocognitive decision types.

	Coordination	***Sensemaking***	***SA***	***Managing Uncertainty and Risk***
Communication	0.0 (0.0)	1.2 (1.8)	0.38 (0.65)	0.0 (0.0)
Enemy Status/Action	2.2 (0.2)	1.7 (1.5)	2.3 (1.1)	1.6 (1.4)
Game Environment/Objective	0.17 (0.41)	0.69 (0.95)	1.0 (1.6)	1.7 (1.6)
Game Sense	0.8 (1.8)	2.8 (1.9)	1.6 (1.3)	2.0 (1.4)
Teammate Status/Action	1.2 (1.2)	2.2 (1.7)	1.8 (2.4)	1.4 (1.1)

Table 22 Means and standard deviations for CoA type across Macrocognitive type.

	Coordination	**Sensemaking**	**SA**	**Managing Uncertainty and Risk**
Action	0.5 (0.55)	1.1 (0.76)	0.67 (0.5)	0.8 (0.84)
Inaction	0.33 (0.52)	0.31 (0.48)	0.11 (0.33)	0.4 (0.55)
Planning	0.5 (0.55)	0.31 (0.85)	0.33 (0.5)	0.6 (0.89)

7 Chapter 7: Discussion

The Tracer Method has been demonstrated to be useful in validating crossover information across methodologies, providing different information between methodologies, and providing new information with the combination of these sub-methodologies. The percentage of mappable cues is critical because it describes the maximum amount of possible overlap between the two methods, which ended up being 60%. On the surface, the difference in "information" is obvious, but our goal was to go deeper and understand the extent and meaning of the data. The confirmed rate of cues within the 20 second decision window provides an indication of overlapping or "redundant" information from the CTA and ET. Of the 170 mappable cues, 59% of the cues were identified validated by ET. Eye tracking itself would not be able to track or identify specific sound cues. The rest of the unmappable cues that were not sound-based cues, further articulates the difference between the two methodologies and the information that they provide. Cues such as, "Moira about to ult," "Opposing team had low range," or "Junkrat has to cover distance" are all cues that articulate awareness or a sense relating to the game that most likely comes from expertise and their ability to predict and anticipate (Klein et al., 2011). There is no systematic way to map these cues with eye gaze, thus, CTA provides insight on these types of cues that we could not get from eye tracking. Eye tracking, on the other hand, provides behavioral insight into fixation counts, durations, and order, which CTA is not able to provide.

The purpose of combining these two methods is to assess whether or not the combination is useful. Table 23 described the benefit of The Tracer Method relative to the combination of process tracing and ET. ET relies heavily on the context of the problem or question, because without this information, the ET has little to no meaning. The CTA or interview method grounds the eye tracking into scope. There may be measures of ET that are more useful (i.e. fixations) than others (i.e. saccades) for different research questions. Another benefit to adding ET is its flexibility. It is able to be incorporated into many different environments, and the output remains generally consistent. CTA, on the other hand, may provide similar output across different sub-methods, but can depend highly on context. In other words, not every CTA method will produce the same information, and not every CTA method is meant to be used in every type of environment or scenario. This research is different from that of Cooke (2010) and Elling et al. (2012) for several reasons: 1. The environment complexity, 2. The CTA methodology, and 3. The focus of the research. The dynamic complexity of a video game, especially a first person shooter, is much greater than navigating a website. Thus, a concurrent think aloud protocol would not have been useful, as it would have disturbed the players' information process as well as communication with teammates. Even a retrospective think aloud would not have been sufficient because the questions are not as guided or established like the CDM probes have been. Finally, the focus of the research from Cooke (2010) and Elling et al. (2011;2012) was to establish the combination as a usability method for static environments, and to categorize the different types of verbalizations and behaviors during silences, in which they are unable to necessarily speculate confidently on what cognitive processes are occurring. Using CDM provides direct insight into these cognitive processes and does not require unjustified speculation. It is also difficult to tie eye tracking movements to decision making or learning processes

(Bednarik & Tukiainen, 2006; Kim et al., 2012), but it can be used for validation of behavioral utterances or processes.

Table 23 Benefits to using The Tracer Method.

Method	**Process Tracing + ET**	**The Tracer Method (CTA + ET)**
Data Window	Full Task (e.g. 30 mins)	Sample (e.g. 40-80 seconds)
Unit of measurement	Thought units (e.g. words, phrases)	Critical decisions, cues, and CoAs
Calculated overlap	ET confirmed thought units (e.g. locations, cues)/Total thought units	ET confirmed cues/Total CTA cues
Depth of data	Shallow	Deep
Level of intrusion	Concurrent= A lot Retrospective= None	None
Tested environments	Static (e.g. websites)	Dynamic (e.g. scene of a fire)
Scalable	Limited	Yes

Regardless of the goal or problem space, The Tracer Method can support both validation and exploration of current and future adjustments to training or developing assessment methodologies. By analyzing the data by each individual sub-method, we are able to understand the contribution of the combination. If one method individually is greater than the contribution of both combined, then the methodology should not be further refined. Our data suggests, however, that these methodologies can be more beneficial when used together than when used separately.

7.1 Limitations

There are several limitations to the conducted study involving equipment, personnel, and methodology. Participants' Gaze Sample percentage (dividing the number of usable samples by number of attempts) ranged from 56% to 95%, meaning that just over 50% of the time, one eye or both eyes were found for half of the recording duration. For 1-2 participants, their gaze data was non-existent for a portion (ranging in 5-15 seconds) of their decision segment. For all of the participants, the recorded framerate was less than 15 ms, despite the eye tracker sampling at a much higher rate. This low framerate made video coding more difficult, as the video was jumpy and choppy. To remedy this, fixations in the PVAL were identified by the immediate location.

Another important limitation to note is both in sample size and gender distribution. Seventeen participants may be perceived as low, however, it is not uncommon to have less than 20 participants in CTA research. Specifically using CDM, participants are typically being interviewed for 1.5-2 hours. In our study, we spent at least 3 hours with each participant, which is more than typical. Of the seventeen participants, only one identified as being female, which presents a large gender imbalance. At our university, gender is largely imbalanced at 75% males and 25% females. Our sample gender discrepancy was not by design, and we did not intend on exploring any gender differences as there has been no past research indicating that there should be. Overwatch's player base is known to have a more diverse distribution across gender, and the game itself has been described as celebrating diversity across gender and ethnicity.

By restricting decision lengths to 20 seconds for standardization reasons and to ensure that the decision windows did not overlap, we cannot address the temporal element of each decision type. Strategic decisions are expected to have a larger temporal quality, thus, by restricting the interval, we may be missing data. This is a difficult problem to solve, as we chose to control for length of decision. Another issue stemming from the time frame of the decision was that participants were not asked to specify the exact start and end of a decision point (when their information or planning began and execution finished). This will be implemented in future studies using The Tracer Method.

Specifically relating to ET, the minimum fixation duration was specified at 60ms, which is the default in Tobii Pro Studio. There has been no research done in complex environments, such as video games, to determine the appropriate minimum fixation duration. The most applicable research was done by Backs & Walrath (1992) investigating eye movement across multiple symbolic displays. Their minimum fixation duration was set to 83ms, and their findings indicate that people required less fixation (or dwell) duration when conducting exhaustive search, often in a sequential or linear fashion, for a symbol or element of a display. Thus, during this search, people may only need to process 1-2 features of an element to know its status. For example, color or lighting may indicate whether something is on or off. Although the 60ms minimum fixation duration may be concerning when investigating decision making in games, our findings suggest that the minimum fixation duration may differ across AOI or element type. For example, the PVAL had the highest average fixation duration (~220 ms), as it is the most complex and dynamic AOI. Previous research suggests that tasks such as reading and other deeper processing tasks require between 200-300ms plus processing time (Salthouse & Ellis, 1980). Future analyses should explore the minimum fixation duration across AOIs or use a greater number, such as 100ms, suggested by Jacob & Karm (2003).

7.2 Applications of Findings

7.2.1 Training

The efficiency and ease of this methodology is applicable to evaluating and development of training methods. Incorporating ET with the current CTA methodologies will inform training in several ways: establishes some validation of interview output, provides new insights into behaviors not reported in the interview, and ensures the focus of the current training is in line with the results from expert interviews. Eye tracking keeps the CTA grounded in “reality” which is sometimes difficult. People can tend to think in the “ideal” instead of their real behavior, which can affect the CTA data. This is not a fault of the method. Eye tracking helps maintain the scope and accuracy of the CTA data. Obviously, eye tracking is not able to completely validate the data, as this study demonstrates that not all information is mappable to eye tracking. However, this does not limit ET’s contribution; Utilizing a bottom up approach, as used in this study, can provide additional insights for the data output. If we assume that the not reported data (found in ET) is due to expert automation of the information processing stage, it will help experts break down the process to inform training. These processes may stem from “weak signals” that the current training regime does not support. The output of The Tracer Method helps ensure that the current training process is in line with and supporting the current user or operator population. If the results from the interview suggest specific visualization patterns or cognitive strategies that are not supported by current training, then the method would call for a restructuring of the training, and may suggest areas of improvement.

7.2.2 Game Research and Design

In addition to helping evaluate and develop training methodologies, *The Tracer Method* can help game designers understand the dynamics and flow of their game, from the perspective of an expert player's mindset, strategies, and behaviors. This data suggests that in *Overwatch*, the roles are balanced in terms of decision making load. Each role performs cognitively difficult tasks in every game, as demonstrated by the Task Diagram and CDM outputs. It is not necessarily a bad thing that there aren't many differences across roles in terms of type of decisions. This articulates that the cognitive and execution requirements across roles are well-balanced. One role is not making most of the decisions. These findings indicate the richness of *Overwatch* data, and that we can elicit different types of incidents and decisions from players. The roles are interdependent and support each other during gameplay. On a larger scale, this CTA method allows a general understanding of the types of challenges that players face most frequently and what specifically makes them difficult. Game designers can use this method to unpack how players' interact in their game and make decisions about how to adjust abilities, rules, or other mechanics to better balance the responsibilities across players. This can be useful to designers who have established and successful games who want to make it better, or designers whose games are in early stages of development and need guidance on how to structure the environment or mechanics to support player decision making.

The Tracer Method could also inform other elements of game design, such as user interface design. In *Overwatch*, players are able to gauge the health of their teammates in many ways: looking at their health bar, seeing a visual critical health indicator, hearing a teammate's hero call out "*I need healing"*, hearing a teammate's hero grunt from taking damage or comment on their status (i.e. Mercy saying, "I'm under attack"). A player may not need to use all of these cues to know that Mercy's health is low, but it is important to understand which cues are useful, or which events require additional cues of information. Another example of how the output of the Tracer Method could support game design is unpacking expert strategies and "mental shortcuts" for how they make quick and efficient decisions. Recently, Blizzard has incorporated the ability for players to easily access each team's individual stats regarding most played heroes and their rank. This allows players to potentially make more informed and strategic decisions regarding team composition and distribution of team responsibilities. There is a limited time for both teams to prepare before a match (less than a minute), thus, having this information more accessible versus requiring players to individually check each person's profile encourages them to make more "expert-like" decisions. The benefit and application of the Tracer Method to support game design is not limited to *Overwatch.* We believe that the evidence for support in game design will be consistent, if not stronger in other genres of games and games at different production stages.

7.2.3 Esports

Within the context of *Overwatch*, *The Tracer Method* has elicited unique and intense cognitive information that Esports teams and coaches seek to improve upon. For example, understanding the decision types and macrocognitive functions of different roles can help teams restructure how they "shot call" (or make callouts and commands) in game. Obviously there are individual differences across pro players, where some players may tend to be more vocal than others. This method supports a more efficient way of delegating responsibilities based on the tendencies of different roles and the type of information they tend to focus on the most. Through focusing on the moments with the most action and interaction (where players are making decisions), feedback is actionable. You know what players are using to make decisions and how they are using it. For example, some eSports teams have their supports serve as the shot caller because their positioning typically allows them to see everything in front of them. This may or may not be supported if the Tracer Method was used on their teams. It may depend on the type of decision (Strategic, Operational, Tactical) in addition to role, as supported by the interaction between decision type and role on number of communication cues. For other types of cues (Teammate/Enemy status or action, or Game sense), having a specific role be a shot caller is not supported by these findings. It may be difficult to determine the ideal shot caller without operationalizing the criteria for what he/she should do or information that they should have.

Our sample data did elicit information from two players on our university's TESPA collegiate Esports team. These individuals are not professional players, but they both compete and perform at a level that is above the average *Overwatch* player. From our data, we can speculate on how using The Tracer Method can support post-assessment of pro games, changes in team dynamics or responsibility delegations, or even player recruitment. Currently, players are evaluated using a combination of their rank (how good are they compared to all other players), their average statistics across games (includes relevant data per role; i.e. eliminations and damage for DPS players), how consistent their gameplay is, and perhaps the depth of their hero pool (how many heroes are you confident in playing in competitive settings). Even with decision making being a major part of the game, there is no systematic way to assess players on it or provide insights to players on how to improve it, until The Tracer Method.

7.3 Future Work

The combination of these two methodologies also provides encouragement for follow up interviews with participants. With the new insights on unreported cues, the participant can be prompted to describe their visual strategies or reasons behind utilizing certain cues that they did not verbalize. They may reveal that these cues have become automatized and they are not actively processing it or that they don't remember looking at it. Another insight for future studies is more closely evaluating which CTA methodology fits the best within the context. For example, within the context of a video game, a simulation interview may have fit better. As this study was being designed, it was difficult to determine this as there has been no research of any other CTA method besides verbal protocols or process tracing methods. CDM is the most similar to retrospective verbal protocol in terms of its temporal and general interview focuses. More research needs to be done using different CTA techniques in video games to assess the most effective method to elicit information.

As previously mentioned, there has been limited research done using both CTA and ET. This is the first demonstration of it in a complex (both cognitively and visually) environment. We intend on investigating The Tracer Method in different types of games (e.g. RTS, MOBA, or MMORPG). We are currently in the process of evaluating The Tracer Method in Civilization V, a turn based strategy game and multiple MOBAs. Future research is necessary to fully solidify this methodology and establish standard protocols for it. More research also needs to be done in more complex environments outside of video games (e.g. military field tests, medical training, etc.). Testing this methodology in a range of fields will hopefully demonstrate its generalizability and the benefits of using this method in multiple domains.

8 Chapter 8: Conclusion

This paper presents the first documentation of a combined methodology made up of Cognitive Task Analysis and ET meant for dynamic, complicated, and changing environments: The Tracer Method. The purpose of this method isn't to discredit the use of them separately, but to provide practitioners and researchers a lightweight, new, flexible, and systematic method to scope their projects and validate their data. This research contributes to the Human Factors domain by taking these methodologies in a new direction, and providing a novel method to execute interdisciplinary research. The added value of the Tracer Method is two-fold: It yields conceptually better information, and is a more efficient method as it reduces workload and is deeply focused on the important parts of the data. We are not the first to suggest that these two methods are complimentary, but we are the first to demonstrate how CTA can increase context and reduce research workload. The Tracer Method is by no means perfect, but similar to CTA and ET methods, it takes time and sufficient amounts of testing to refine. We support the alliance of CTA and ET, and articulate that these two methodologies benefit in duo versus going solo. The world could always use more methods!

9 References

Alkan, S., & Cagiltay, K. (2007). Studying computer game learning experience through eye tracking. *British Journal of Educational Technology*, *38*(3), 538-542.

Almeida, S. D. J. (2009). *Augmenting video game development with eye movement analysis* (Master's thesis, Universidade de Aveiro).

Almeida, S., Veloso, A., Roque, L., & Mealha, Ó. (2011). The eyes and games: A survey of visual attention and eye tracking input in video games. In *Proceedings of the SBGames*, 1-10.

Anders, G. (2001). Pilot’s attention allocation during approach and landing—eye- and head-tracking research in an A330 full flight simulator. In *Proceedings of the 11th International Symposium on Aviation Psychology*.

Annett, J. (1996). Recent developments in hierarchical task analysis. *Contemporary ergonomics*, 263-268.

Bednarik, R., & Tukiainen, M. (2006, March). An eye-tracking methodology for characterizing program comprehension processes. In *Proceedings of the 2006 symposium on Eye tracking research & applications* (125-132). ACM.

Bolstad, C. A., Riley, J. M., Jones, D. G., & Endsley, M. R. (2002, September). Using goal directed task analysis with Army brigade officer teams. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 46, No. 3, pp. 472-476). Sage CA: Los Angeles, CA: SAGE Publications.

Boyle, E., Van Rosmalen, P., MacArthur, E., Connolly, T., Hainey, T., Johnston, B., & Manea, M. (2012, October). Cognitive task analysis (CTA) in the continuing/higher education methods using games (CHERMUG) Project. In *6th European Conference on Games Based Learning* (63).

Brezovic, C. P., Klein, G. A., & Thordsen, M. (1987). Decision making in armored platoon command (KATR-858 [B]-87-05F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327 for the US Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Buswell, G. T. (1937). *How adults read*. Chicago, Ill: University of Chicago.

Calderwood, R., Crandall, B. W., & Klein, G. A. (1987). Expert and novice fire ground command decisions (No. KA-TR-858 (D)-87-03F). KLEIN ASSOCIATES INC YELLOW SPRINGS OH.

Cooke, L. (2010). Assessing concurrent think-aloud protocol as a usability test method: A technical communication approach. *IEEE Transactions on Professional Communication*, *53*(3), 202-215.

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, *41*(6), 801–849.

Cooke, L., & Cuddihy, E. (2005 Using eye tracking to address limitations in think-aloud protocol. In *Professional Communication Conference, 2005. IPCC 2005. Proceedings. International* (653-658).

Crandall, B. W. (1989). A comparative study of think-aloud and critical decision knowledge elicitation methods. *ACM SIGART Bulletin*, (108), 144-146.

Crandall, B., and R. Calderwood 1989 Clinical assessment skills of experienced neonatal intensive care nurses. National Center for Nursing Research, National Institutes of Health. Fairborn, OH: Klein Associates.

Crandall B., Gamblian V. (1991). Guide to early sepsis assessment in the NICU. Fairborn, OH: Klein Associates Inc. Instruction manual prepared for the Ohio Department of Development under the Ohio SBIR Bridge Grant program

Crandall, B., & Getchell-Reiter, K. (1993). Critical decision method: a technique for eliciting concrete assessment indicators from the intuition of NICU nurses. Advances in Nursing Science, 16(1), 42-51.

Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. MIT Press.

Crandall, B., Klein, G, Militello, L. G., & Wolf, S. P. (1994). Tools for applied cognitive task analysis (Contract Summary report on N66001-94-C-7008). Fairborn, OH: Klein Associates.

Dearborn, W. F. (1906). *The psychology of reading*.

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 455–470.

Duchowski, A. T. (2007). Eye tracking methodology. *Theory and Practice, 328*.

Elling, S., Lentz, L., & De Jong, M. (2012). Combining concurrent think-aloud protocols and eye-tracking observations: An analysis of verbalizations and silences. *IEEE Transactions on Professional Communication, 55*(3), 206–220.

Elling, S., Lentz, L., & de Jong, M. (2011). Retrospective think-aloud method: Using eye movements as an extra cue for participants' verbalizations. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (1161-1170).

El-Nasr, M. S., & Yan, S. (2006). Visual attention in 3D video games. In *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology* (22). ACM.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*(4), 327.

Gallagher, P. S., & Prestwich, S. (2013). Cognitive Task Analysis: Analyzing the Cognition of Gameplay and Game Design. Paper presented at *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. Orlando, FL.

Gidlöf, K., Wallin, A., Dewhurst, R., & Holmqvist, K. (2013). Using eye tracking to trace a cognitive process: Gaze behaviour during decision making in a natural environment. *Journal of Eye Movement Research*, *6*(1).

Glaholt, M. G., & Reingold, E. M. (2011). Eye movement monitoring as a process tracing methodology in decision making research. *Journal of Neuroscience, Psychology, and Economics*, *4*(2), 125.

Goldberg, J. H., & Wichansky, A. M. (2003). Eye tracking in usability evaluation: A practitioner's guide. In *The Mind's Eye* (493-516).

Granka, L. A., Joachims, T., & Gay, G. (2004, July). Eye-tracking analysis of user behavior in WWW search. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (478-479). ACM.

Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (1253–1262). ACM.

Hoffman, R. R. (1987). The problem of extracting the knowledge of experts from the perspective of experimental psychology. AI magazine, *8*(2), 53.

Hoffman, R. R., Crandall, B., & Shadbolt, N. (1998). Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors*, *40*(2), 254–276.

Hoffman, R. R., & Militello, L. G. (2012). Perspectives on cognitive task analysis: Historical origins and modern communities of practice. *Psychology Press.*

Hoffman, R. R., Shadbolt, N. R., Burton, A. M., & Klein, G. (1995). Eliciting knowledge from experts: A methodological analysis. *Organizational behavior and human decision processes, 62*(2), 129-158.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011a). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011b). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Horn, B., Cooper, S., & Deterding, S. (2017, October). Adapting Cognitive Task Analysis to Elicit the Skill Chain of a Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (277-289). ACM.

Huey, E. B. (1908). *The psychology and pedagogy of reading*. The Macmillan Company.

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, *2*(3), 4.

Johansen, S. A., Nørgaard, M., & Soerensen, J. R. (2008). Can eye tracking boost usability evaluation of computer games?. In *Conference on Human Factors in Computing Systems, CHI 2008*.

Jonassen, D. H., Tessmer, M., & Hannum, W. H. (1998*). Task analysis methods for instructional design*. Routledge.

Jönsson, E. (2005). If looks could kill–an evaluation of eye tracking in computer games. *Unpublished Master's Thesis, Royal Institute of Technology (KTH), Stockholm, Sweden*.

Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009, September). Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th international conference on* (2106-2113). IEEE.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, *87*(4)

Kaakinen, J. K., & Hyönä, J. (2010). Task effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6)

Kaempf, G. L., Wolf, S., Thordsen, M. L., & Klein, G. (l992). Decision making in the AEGIS combat information center. Fairborn, OH: Klein Associates Inc. Prepared under contract N66001-90-C-6023 for the Naval Command, Control and Ocean Surveillance Center, San Diego, CA.

Kaufman, A. E., Bandopadhay, A., & Shaviv, B. D. (1993, October). An eye tracking computer user interface. In *Virtual Reality, 1993. Proceedings., IEEE 1993 Symposium on Research Frontiers in* (120-121).

Keefe, R. S., Silverman, J. M., Mohs, R. C., Siever, L. J., Harvey, P. D., Friedman, L., ... & Davis, K. L. (1997). Eye tracking, attention, and schizotypal symptoms in nonpsychotic relatives of patients with schizophrenia. *Archives of general psychiatry*, 54(2), 169-176.

Kenny, A., Koesling, H., Delaney, D., McLoone, S., & Ward, T. (2005). A preliminary investigation into eye gaze data in a first person shooter game. In: *19th European Conference on Modelling and Simulation (ECMS 2005)*

Klein, G. A., Calderwood, R., & Macgregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on systems, man, and cybernetics*, *19*(3), 462-472.

Klein, G. A., & Thordsen, M. L. (1991). Representing cockpit crew decision making. In *6th International Symposium on Aviation Psychology*.

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent systems*, *18*(3), 81-85.

Klinger, D. W., Andriole, S. J., Militello, L. G., Adelman, L., & Klein, G. (1993). *Designing for performance: A cognitive systems engineering approach to modifying an AWACS human computer interface*. KLEIN ASSOCIATES INC FAIRBORN OH.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*(6483), 742-744.

Law, B., Atkins, M. S., Kirkpatrick, A. E., & Lomax, A. J. (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* (41–48). ACM.

Leyba, J., & Malcolm, J. (2004). Eye tracking as an aiming device in a computer game. *Course Work (CPSC 412/612 Eye Tracking Methodology and Applications by A. Duchowski), 14*. Clemson University.

Mello-Thoms, C., Nodine, C. F., & Kundel, H. L. (2002). What attracts the eye to the location of missed and reported breast cancers? In *Proceedings of the 2002 symposium on Eye tracking research & applications* (111–117). ACM.

Militello, L. G., & Hutton, R. J. (1998). Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, *41*(11), 1618–1641.

Mislevy, R. J., Steinberg, L. S., Breyer, F. J., Almond, R. G., & Johnson, L. (1999). A cognitive task analysis with implications for designing simulation-based performance assessment. *Computers in Human Behavior*, *15*(3), 335–374.

Mullinix, G., Gray, O., Colado, J., Veinott, E., Leonard, J., Papautsky, E. L., Flach, J. (2013). Heuristica: Designing a serious game for improving decision making. In *2013 IEEE International Games Innovation Conference (IGIC)* (250–255).

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231.

Nivvedan, S. (2014) *Literature Survey on Eye-Tracking.*

Olsen, A., Smolentzov, L., & Strandvall, T. (2010, September). Comparing different eye tracking cues when using the retrospective think aloud method in usability testing. *Proceedings of the 24th BCS Interaction Specialist Group Conference* (pp. 45–53). British Computer Society.

Bailey, Hugh. Open Broadcaster Software. [Software]. Open Source

Ottati, W. L., Hickox, J. C., & Richter, J. (1999, September). Eye scan patterns of experienced and novice pilots during visual flight rules (VFR) navigation. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 43, No. 1, pp. 66-70). Sage CA: Los Angeles, CA: SAGE Publications.

Overwatch. (2016) [Video Game]. Irvine, California: Blizzard Entertainment.

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. In C. Ghaoui (Ed.), Encyclopedia of human-computer interaction. (211-219). Pennsylvania: Idea Group Inc.

Randel, J. M., Pugh, H. L., & Reed, S. K. (1996). Differences in expert and novice situation awareness in naturalistic decision making. *International Journal of Human-Computer Studies*, 45(5), 579-597.

Recarte, M. A., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied*, *6*(1)

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, 51(1), 1-41.

Renkewitz, F., & Jahn, G. (2012). Memory indexing: A novel method for tracing memory processes in complex cognitive tasks. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(6).

Renshaw, T., Stevens, R., & Denton, P. D. (2009). Towards understanding engagement in games: An eye-tracking study. *On the Horizon*, *17*(4), 408–420.

Rhenius, D., & Deffner, G. (1990, October). Evaluation of concurrent thinking aloud using eye-tracking data. In *Proceedings of the human factors society annual meeting* (Vol. 34, No. 17, pp. 1265-1269). Sage CA: Los Angeles, CA: SAGE Publications.

Roose, K. M., & Veinott, E. S. (2017, October). Roller Coaster Park Manager by Day Problem Solver by Night: Effect of Video Game Play on Problem Solving. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play* (pp. 277-282). ACM.

Salvucci, D. D., & Goldberg, J. H. (2000, November). Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the 2000 symposium on Eye tracking research & applications (pp. 71-78). ACM.

Schaafstal, A., Schraagen, J. M., & Van Berl, M. (2000). Cognitive task analysis and innovation of training: The case of structured troubleshooting. *Human Factors, 42*(1), 75–86.

Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (2000). *Cognitive task analysis*. Psychology Press.

Schweickert, R., Fisher, D. L., & Proctor, R. W. (2003). Steps toward building mathematical and computer models from cognitive task analyses. *Human factors*, *45*(1), 77-103.

Seagull, F. J., Xiao, Y., MacKenzie, C. F., Jaberi, M., & Dutton, R. P. (1999, September). Monitoring behavior: A pilot study using an ambulatory eye-tracker in surgical operating rooms. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 43, No. 15, pp. 850-854). Sage CA: Los Angeles, CA: SAGE Publications.

Seagull, F. J., & Xiao, Y. (2001a). Using eye-tracking video data to augment knowledge elicitation in cognitive task analysis. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 45, pp. 400–403). SAGE Publications Sage CA: Los Angeles, CA.

Seagull, F. J., & Xiao, Y. (2001b). Using eye-tracking video data to augment knowledge elicitation in cognitive task analysis. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 45, pp. 400–403). SAGE Publications Sage CA: Los Angeles, CA.

Shute, V. J., & Kim, Y. J. (2011). Does playing the World of Goo facilitate learning. *Design Research on Learning and Thinking in Educational Settings: Enhancing Intellectual Growth and Functioning*, 359–387.

Smith, J. D., & Graham, T. C. (2006). Use of eye movements for video game control. In *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology* (p. 20). ACM.

Stanton, N. A. (2006). Hierarchical task analysis: Developments, applications, and extensions. Applied ergonomics, 37(1), 55-79.

United States of America, USAF College of Aerospace Doctrine, Research and Education. (1997). *Three Levels of Wa r*(Vol. 1). AL: Air University Press.

Van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior*, *25*(2), 325–331.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. (2005). Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology*, *19*(2), 205-221.

Veinott, E. S., Perleman, B., Polander, E., Leonard, J., Berry, G., Catrambone, R., Lemaster, L. (2014). Is more information better? Examining the effects of visual and cognitive fidelity on learning in a serious video game. In *2014 IEEE Games Media Entertainment* (pp. 1–6). https://doi.org/10.1109/GEM.2014.7048105

Zsambok, C. E. (2014). Naturalistic decision making: where are we now?. In *Naturalistic decision making* (pp. 23-36). Psychology Press.

Appendix A. Critical Decision Method Interview Guides

A.1 CDM Pre

Interview Guide for Interview 1 (CDM PRE)
Overwatch Beta: October 27, 2015, **Overwatch Release:** May 24th, 2016

Introduction: Thank you again for participating. Just because I cannot write everything down, are you okay if I record the interview?

I am interested in the decisions that you have made and the decision strategies that you have developed in video games, specifically in Overwatch. I will be asking you to talk about various games that were difficult, complicated, and had critical decisions throughout the game.

Warm up (5 minutes MAX):

1. How many months have you been playing Overwatch?
2. How often do you play per week? What modes?
3. Who do you play with?
4. What roles do you play?
5. What heroes do you play?
6. What are your best roles?
7. What are your best heroes?

Task diagram: As a "x" what are 3 of your cognitively difficult tasks during the game?
For each task:

1. Why is this cognitively difficult?
2. What are your strategies for managing "y?" (With "y" being the task)

Sweep 1: Identify incidents

1. I am interested in hearing about examples in specific games of Overwatch when you made a really bad decision as "x."
    a. How often does this situation happen with you in Overwatch games?
    b. Can you tell me about examples in specific games of Overwatch when you made a really good decision as "x."
2. What about a game in which something unexpected happened? Can you set the scene and take us back to that game?
    - How often does this situation happen with you in Overwatch games?
3. **ALT**: I am interested in your decision making process during games when you play "x." Can you describe the game environment and set the scene for when there was a critical point in a specific game when you had to make a difficult decision?
    - **If they have difficulty, ask them:** You have played a lot of Overwatch. Tell me about one of the most recent games of Overwatch

Sweep 2: Develop a timeline

1. Based on (*refer to chosen incident here*) that you just described, let's make a timeline of the event that you just described. (Repeat the details back to them) Do I have the details right so far?
2. What events happened before the critical incident?
3. Can you tell me some of the events that occurred as a result of the critical decision (what came after)?

Sweep 3: Deepening probes

1. While you were (*refer to decision here)*, what were you noticing in that game?
2. What were you seeing in that game?
   - Other players, visuals
3. What were you hearing in that game?
   - Sounds, music
4. What information did you use in making this decision?
5. What let you know that this was the right thing to do at this point in that game?
6. What other courses of action were considered? How was this option chosen and the others rejected?
7. Did you imagine the possible consequences of the decision that you ended up making? Did you imagine the events proactively and how they would unfold?

Sweep 4: What if scenarios

1. If a novice player had been in your shoes at this particular point in the incident, what would they have done?
   - What mistakes would they have made and why?
2. Would they have noticed what you noticed?
3. Would they have known to (*refer to decision here)*?
4. *Based on the information that the participant gives, some what-if questions might include:*
   - *What if you didn't have "x" information? How would that have affected your decision process?*
   - *What if "x" happened? How would that have affected your decision process?*

A.2 CDM Post

Interview Guide for Interview 2 (CDM POST)

Background Info: Age, Gender, Major

Sweep 1: (Before showing them the footage)

1. In the "x" game that you just played, what were your critical decisions in the game?
   a. Rephrased: In the game you just played…

i. "what were some decisions you made that mattered?"
ii. "what were some important decisions you made?"

Answer: Critical decision 1 (**D1**), Critical Decision 2 (**D2**), Critical Decision 3 (**D3**)…

Sweep 2: (Before showing them the footage) Use Sticky Notes*

Instructions: I would like to construct a timeline of decisions you just mentioned to get general timeline and events leading up to each decision.

1. In what order did these decisions happen?
2. What were 1-2 events that led up to "x"? (Decisions 1, 2, 3, etc.)

(Repeat the summary of each decision and events)

3. Are the details correct? (over all decisions)
4. Any other critical events missing from this timeline? (over all decisions)

***Sweep 3:* (Beginning with the first critical incident in the footage- show them the footage first then ask questions; using each of the sweep 3 then 4 probes for Decision 1 first, then Decision 2… in order by incident)**

1. Goals: What was your goal or objective at this point in time (e.g., D1)?
- Rephrase: What was **most important** to accomplish?

2. While (refer to the critical incident D1) was unfolding, what were you **paying attention** to or noticing to at that point?
- **Rephrase:** What cues did you have?
- What were you **seeing**?
- What were you **hearing**?

3. While (refer to the critical incident) was unfolding, **what information** did you use in making your decision or judgment?

a. **Follow up:** Where did you get the information from?
- Rephrase: Where did you gather/locate information from to make your decision?
  - **Follow up:** What in "x" or what info from "x" would you get?
  - **Follow up:** How did you get the information?
  - **Follow up:** What did you do with the information?

4. What were your options at this point (for this decision)?
- **Rephrase:** What other courses of action were considered?
  - **How** did you choose this option and reject others?
  - **Why** did you choose this option and not others?
  - **Did** you imagine the possible consequences of this action?

Make note of other probes that you ask and make sure that you ask those for the other critical incidents as well

Sweep 4: (Use each of these probes for each incident)

1. If a novice player had been in your shoes at this particular point in the incident, what would they have done?
   - What mistakes would they have made and why?

2. Would they have noticed what you noticed?
   - Would they have known to (refer to decision here)?

3. Based on the information that the participant gives, some what-if questions might include:

   - **What if** you didn't have "x" information? How would that have affected your decision process during this incident?
   - **What if "**x" happened? How would that have affected your decision process during this incident?

4. How would you train a novice player to improve their gameplay?

   - What would you tell them to **look out for**?
   - What would you tell them to **look at**?
   - What would you tell them to **listen for**?

Appendix B. The Tracer Method Tutorial and Best Practices

B.1 Tutorial

This section provides an overarching tutorial on how to use the Tracer Method. The steps include:

1. **Establish appropriateness of The Tracer Method use.** Is The Tracer Method appropriate to evaluate your research question? Is your intended environment appropriate to use the Tracer Method in? If this method will not help you address your research question, you may need to find another method.

2. **Determine the best CTA method to use.** What CTA method would provide the best output to evaluate your research question?

3. **Determine the specific environment to use the Tracer Method in.** Is this environment possible to use both parts of the Tracer Method? Will you get good data using both methodologies? For example, in outdoor environments, eye tracking may not provide as accurate data as it would indoors.

4. **Develop and scope your interview guide.** What are the main steps of the chosen CTA method? Are there any established or validated questions or probes (e.g. mental model probe from CDM) for the specific CTA method you chose? Use these as guides or use the exact questions from the research.

5. **Establish the procedure and pilot it (1-2 people).** This helps flush out any issues with the interview guide (adding, deleting, etc.). It also helps the interviewers practice, as CTA interviews are not easy to execute.

6. **Create a data management plan.** Important to keep methodology data separate and label updated data files.

7. **Execute the study and note any anomalies in data collection.** This insight shapes future studies and helps proceeding coding

8. **Transfer the CTA data systematically into files for analysis.** If you intend on running different types of statistical analyses (e.g. MANOVA or PCA), ensure that there are multiple formats of the same data that will allow you to correctly perform the analyses. (R Studio is an open source program that is great for data editing and analysis.)

9. **Establish the type and extent of subjective coding that needs to be done based on the data.** Determine who your raters are. If you need raters with specific domain experience, you may need to contact them before data analysis occurs.

10. **Analyze each method's data separately. Try not to "bleed" observations from one method into the other quite yet.** This is ensuring that the data from each methodology is consistent within itself.

11. **Analyze the data together. Look for connections, patterns, anomalies, and gaps. There will be information connecting the two methodologies. This is the information to focus on (See Chapter 5.5.2), so use both perspectives to describe what the new information is telling you. Use overlapping information for cross validation if it answers your research question.** Schedule follow up-interviews if possible for data that needs to be further unpacked.

12. **Evaluate your hypotheses based on the data using The Tracer Method.** Articulate your findings to the academic community.

B.2 Best Practices

From this initial demonstration and evaluation of The Tracer Method, we have created a list of best practices to help researchers and practitioners use this method based on what we have learned.

1. *Evaluate the problem space and identify your research questions:* As previously stated, CTA can be highly context dependent and the method that you choose can elicit different information than another method. We chose CDM for this demonstration because our research focus was on the critical elements and processes of decision making within a specific context. CDM elicits specific information regarding various elements of decision making.

2. *Understand the type or scope of information that you can elicit from your task or environment:* We chose to use both a Task Diagram and a CDM-pre interview to unpack what type of data we would get within the context of *Overwatch*, as there was no prior research on it. If there has been research using CTA, refer to the outcome of that research to help scope your project and determine the correct CTA method.

3. *Determine which eye tracking metrics fit most closely with your research questions:* If your research focuses primarily on the effect of integrating a new technology into an interface or space on the temporal performance (e.g. reaction time or time to detection) of an operator, time to first fixation or fixation order

may be the most beneficial. If you intend on investigating the importance or use of a feature on your website, fixation counts or durations may fit better.

4. *Continuously ask participants to specify what part of the environment or where they are getting their knowledge or cues from:* Even in Sweep 3 of the CDM interview, participants did not always specify where or how they were getting the information that they articulated in the interview. This can be critical in determining whether or not a fixation on an entity allowed them to elicit that information or if it was for another piece of information. Use visual prompting and refer to specific fixations or patterns that they did not articulate in their interview. With CDM specifically, it is not typically part of the protocol to ask about visual strategies or how people manage their attention. As part of a follow-up, once you have passed through the data, revisit it and try to fill in the gaps.

5. *Do not hesitate to ask participants about negative experiences or put them in positions of failure:* With most of the CDM-Pre interviews, participants described games in which they made generally good decisions. If your research is focused on poor decision making or how people error correct, it will be beneficial to elicit information on their decision making, cues, or cognitive processes during failure or when the outcome was bad. CDMs have been used to investigate failures and gaps in decision making in order to avoid that behavior or mishap in the future. Do not shy away from poor decision outcomes or instances of failure.

Appendix C

See supplemental materials for the Fair Use Evaluation of Figures 5, 7, 8, and 15. The base screenshots are all from the game, *Overwatch* (Blizzard, 2016).