GESTURES IN THE VIRTUAL ENVIRONMENT: THE USE OF DEIXIS IN YOUTUBE GAMING TUTORIALS.

by

RACHEL ANNA MANLEY

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

Department of English Language and Applied Linguistics
College of Arts and Law
University of Birmingham
September 2022

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

As natural communicators, humans establish meaning by anchoring their language to the space around them. One method that is used to establish communication is through the use of deictic gestures and expressions. Much of the research to date has focused on use of deictic behaviors in face-to-face interactions within the physical world. However, with the advent of new technology and new ways to communicate, researchers have taken interest in the ways humans interact in the virtual environment. This thesis investigates a small portion of virtual communication by examining the use of deictic behaviors in YouTube gaming tutorials. The goal of the creators of these tutorials is to communicate with their audience, in this research I posit that one of the ways they do so is through deictic behaviors that have parallels to those used in the physical world. In order to examine if this is the case, I analyzed video data through the program ELAN and formulated a unique typology of virtual deictic gestures that have roots in previously established gesture theory.

Dedication

I would like to thank my mother, whose continued support has motivated me to never give up on my goals. Without her encouragement and support I would not be where I am today.

I would also like to thank my best friend who supported me through this long process. Without her help, this achievement would have been far more difficult and painful.

To my father who passed before he could see me graduate. I will always be his Dr. Peanut.

Table of Contents

Chapter 1: Introduction p. 1

Chapter 2: Theoretical Background and Literature Review p. 9

Section 2.1: Computer Games and YouTube p. 10

Section 2.2: The Virtual Environment and The Physical World p. 13

Section 2.3: Gestures p. 14

Section 2.4: Deictics p. 23

Section 2.5: Speech Acts and the Co-occurrence of Deictic Behaviors p. 32

Section 2.6: Spatial Deixis p. 35

Section 2.7: Clark's Directing-to and Placing-for p. 39

Chapter 3: Methods p. 44

Section 3.1: Potential Games and Selection Criteria p. 44

Section 3.2: Mixed Methods p. 58

Section 3.3: Ethics p. 61

Section 3.4: Annotations and ELAN p. 70

Chapter 4: Typology p. 82

Section 4.1: Directing-to p. 82

Section 4.2: Placing-for p. 96

Chapter 5: Results and Discussion p. 108

Section 5.1: Comparison of Deictic Gestures Across All Three Content Creators p. 108

Section 5.2: Speech Acts that Co-occurred with Gestures p. 111

Section 5.3: Gesture Analysis p. 127

Section 5.4: Clark's Temporal Dimension of Directing-to and Placing-for p. 140

Section 5.5: Demonstratives p. 148

Section 5.6: Demonstrative Choices p. 165

Section 5.7: Deictic Gestures as Communicative Devices p. 175

Chapter 6: Conclusion p. 182

References p. 191

Appendix A-C p. 220

Chapter 1: Introduction

When people communicate, they use the space around them in a multitude of ways to establish meaning. The way we use space is “an integral part of the way we think, plan and behave, a central element in the way we shape the very world that constrains and guides our behavior” (Kirsh, 1995, pp. 31-32). One of the ways communication is anchored to the physical environment is through communicative devices such as spatial deictic expressions, including demonstratives and deictic pointing gestures. When a person wants to draw attention to objects in the physical environment, they refer to them by pointing to the item with their finger or placing the item within the sight of their interlocutors (Clark, 2003). For example, if an individual at a restaurant wanted a refill of their drink, they may place the glass at the edge of the table as a way to signal the server to bring more water. What happens, though, when people communicate in virtual environments? How do they tie their utterances to an environment that is not physical?

In mediated contexts, one way for a person to go beyond physical world communication is to interact with virtual tools in such a way as to extend themselves into the virtual environment (see e.g., Spagnolli, Lombard & Gamberini, 2009). In this environment, people cannot always point with their fingers, and there are no physical objects to incorporate into their communicative practice. Instead of manipulating their body or their physical environment they must use the features of the virtual environment and the computer interface to aid in communication. However, to understand how communication occurs and how it is reshaped by the virtual environment, it is first necessary to explain how it works in the physical world.

In the physical world, we tend to use both gesture and speech concurrently because “we cannot interpret what is actually said without interpreting the gestural activity in conjunction with the verbal utterances…” (Nash, 2007, p. 115). This synthesis of verbal utterance and nonverbal gesture is accomplished in part by deixis. Deictic gestures and deictic expressions are the “linguistic features or expressions that relate utterances to the circumstances of space and time in which they occur” (Kendon & Versante, 2003, p. 133). In other words, deixis is where communication meets context: deictic expressions need knowledge of the local communicative context to be interpretable, and gesture often performs a special role in achieving this link to context.

In the virtual world, it is clear that people can use a computer mouse in conjunction with their speech to relate what they are saying to what they have on the screen (Aogain & Reilly, 1990), thereby demonstrating that mouse cursor movements can support deixis in this particular communicative context. But perhaps more so than in the physical world where our hands are often free during face-to-face conversation, the mouse cursor also performs other functions in the virtual environment, such as functional, goal-directed movements that interact with the software, but that are not meant to be communicative. This poses the question as to how the audience knows how to distinguish between when the mouse is being used to merely interact with software, and when it is being used with communicative intention? I attempt to answer this question by investigating how people use deictic gestures and expressions to anchor meaning in virtual environments in order to communicate with others.

I do this by applying the work from the field of gesture research and pragmatics to a new, mediated environment, YouTube gaming tutorials. One of my primary

frameworks is that of Clark (2003), which is the only account that explicitly splits deictic gestures into two categories, directing-to and placing-for, a distinction which relates to how attention interacts with communication. In the physical world, directing-to occurs when someone uses a pointing gesture to direct their interlocutor's attention to the object of reference. Placing-for is established when a person places objects at an indexing site to which both the speaker and the interlocutor have access, that is, it involves the manipulation of the environment with respect to the addressee's attention. The proposal then is that the distinction between directing-to and placing-for exhaustively captures the ways attention plays into communication, either by directing attention, or by manipulating things with respect to attention. Although I believe Clark's work can be applied to the virtual environment, no one else has yet demonstrated that his dichotomy of directing-to and placing-for holds explanatory power in this specific context, nor has his dichotomy been empirically tested in any context. I hypothesize that deictic gestures used in the virtual environment have similarities and differences to those in the physical world, but that his dichotomy is not as clearly separated as he contends. My data will show that at certain times and under certain conditions, both directing-to and placing-for can be applied simultaneously. In other words, the distinction sometimes breaks down.

Deictic gestures are examined in conjunction with the linguistic aspects of deixis that are used in YouTube gaming tutorials. I will focus on spatial deictic expressions, that is, expressions that *need* some form of nonverbal indication to gain meaning, such as "*That* thing over *there*," which is uninterpretable without a concomitant gesture. These expressions are most likely to show interactions with deictic gestures, in contrast

to nominal expressions (“The blue thing that is ball-shaped”), which rely less on context in order to be interpreted successfully. When and in which contexts speakers choose verbal demonstratives, such as “here” and “there” has been examined in the physical world (e.g., Cooperrider, 2016; Glover, 2000; Hanks, 2017), however, the choice of demonstratives used by content creators of tutorials in the virtual environment has received little attention especially regarding its relationship with deictic gestures.

YouTube gaming tutorials are the focus for this thesis, because they are one of the most watched video types on YouTube (Anderson, 2020), and because, as a gamer, I am deeply familiar with this medium. Gaming tutorials provide an environment in which deixis is fostered because it is a medium individuals use to instruct or inform an audience about something that is seen on the screen. Since tutorials are about how to do something through demonstration and explanation, I expected that content creators would reference objects within their videos using deictic expressions and gestures during their instructions. The expected richness of data in this resource, provides ample new material in which to extend the emerging study of spatial communication that has begun to emerge in internet pragmatics (see Chapter 2 for a fuller discussion of the research to date) and allows us to further consider the creators’ usage of the communicative tool of deictic expressions and gestures as a way to anchor meaning within a widely used, and multi-faceted virtual environment.

As communication in virtual environments is an important part of many people’s lives, it is crucial to investigate the communicative tools found in the virtual environment and how individuals use these tools to communicate with others. The virtual environment is becoming ever more pervasive to the modern way of life, and any theory

of deixis is incomplete if it does not account for a type of environment that is now very dominant. Since video game tutorials involve a great deal of reference to objects in the game world, studying them should further expand our knowledge on how deictic expressions and gestures are used in a variety of contexts, through which we can arrive at a more general theoretical framework for spatial deixis that encompasses both the physical and virtual case. Moreover, a look at deictic mouse cursor movements also enriches gesture research: I argue that the mouse cursor actions individuals take in the virtual environment to establish deixis can be considered gestures on par with manual gestures. One way to substantiate this claim is to point out similarities between communicative mouse cursor movements and manual co-speech gestures that are performed in physical contexts (Cooperrider, 2016; Kendon, 2004; Kita, 2003; McNeill, 1992). As I will highlight, there are deep parallels in the precise execution of these behaviors. If there are enough similarities and it can be demonstrated that the concept of gesture can fruitfully be applied to the virtual environment, then one could state that people not only find and create new ways of gesturing within the virtual environment, they also take the characteristics of gestures in the physical world and adapt them to the virtual environment. Thus, my analysis allows us to view ‘gesture’ as a modality-independent concept that fundamentally shares patterns of expression across different modalities.

Finally, this thesis not only contributes to the field of pragmatics and gesture research, but it is also relevant to the gaming and software development community. In the gaming community, gameplay streaming has increased in popularity, however, there is still room to train content creators on successful methods of communicating with their

audience (Sjöblom, Törhönen, Hamari & Macey, 2019). As this thesis analyzes different deictic expressions and gestures and how to indicate in the virtual environment, the results can shed light on ways content creators can use deictic behaviors to effectively communicate and provide clarity in their referencing. An example of this is knowing what types of gestures may be useful in cluttered environments or what gestures remove ambiguity for those with visual impairments (Jun, Seo, Park, Park & Jung, 2021).

With respect to software development, my analysis can inform how we design games and programs that are used in the creation of video game tutorials. Content creators use software programs, such as OBS and XSplit, to help them record their gameplay. Understanding how content creators use the specific affordances provided by the games they play, as well as the software they use to communicate, can inform future software development. Software developers could benefit from knowing what type of deictic gestures are used or needed by their consumers so they can either create new programs or improve on existing ones. This is essential as content created by other players can have an influence on viewers to buy and play the game they are introducing (Macey, Tyrväinen, Pirkkalainen & Hamari, 2020), and content creators may be influenced to choose streaming and video recording programs based on what more completely addresses their need.

Although my focus on deictic gestures deals with YouTube gaming tutorials, the concepts can be applied to other contexts outside of the gaming community, and outside of the context of YouTube. When individuals work in joint activities (see Clark, 1996), they can use collaborative software such as Google Docs or Zoom to share

screens, highlight, and write, and they can use the mouse cursor or their keyboard as part of deictic gestures to direct attention to specific objects and locations on the screen (Clark's directing-to), or place objects on the screen for their viewers (Clark's placing-for). Because deixis is such a crucial part of communication, we may surmise that increasing opportunities for deixis can help, for example, videoconferencing software gain more popularity. In addition to collaboration in a work environment, research such as the one presented in this thesis can contribute to the future development of learning software (e.g., Wende, Bulut, Giese & Anderl, 2020). Deictic gestures can be used to foster communication between instructors and their learners (Yulfi, 2017) as well as be used in a way to establish joint attention in educational settings, such as online language learning, in which it "provides important scaffolding… for the realization of shared intentionality, which is essential for human communication and cognition" (Stevanovic & Monzoni, 2016, p. 30). The YouTube videos that I investigate are tutorials that help players to become better at the games they are interested in. From this perspective, my thesis also looks at online learning environments, seen through the lens of game learning. This means that my results also have implications for research on online education.

A final topic of relevance is that of "affordances" (Gibson, 1979), which can roughly be described as the utility of an object or what it can provide. For example, a chair affords sitting. The same way, different game environments afford different mouse cursor actions, as well as different interactions with the GUI (Graphical User Interface). Therefore, my study also adds to research that looks at how gestures are "environmentally coupled" (Goodwin, 2007). It is possible that different games are

associated with different gestures, depending on the software tools they make available for establishing deixis.

To guide my analysis, I will rely on the following research questions:

1. What behaviors are used to establish deixis in a virtual environment?
2. How are deictic gestures shaped by the affordances of the games and the software interfaces that are being used in selected YouTube tutorials?
3. How do deictic gestures vary in use by different content creators?
4. How do deictic gestures interact with verbal deictic expressions in these tutorials?
5. How are speech acts tied to their context through deictic expressions and gestures?
6. How do the deictic gestures found in the data relate to existing frameworks from gesture research?
7. How do content creators signify that actions within the virtual environment have communicative function?

The chapters of this thesis are as follows: Chapter 2 will form the literature review and discuss the relevant theoretical background; Chapter 3 will discuss my methodology; Chapter 4 discusses the typology of virtual gestures that I have created on the basis of my data analysis; Chapter 5 will provide and discuss quantitative and qualitative data to show how the gesture types from my typology are used by different content creators; finally, Chapter 6 will present the conclusion of this research.

Chapter 2: Theoretical Background and Literature Review

This chapter outlines the background, concepts, theories, and terminology that is relevant for my thesis. The intent of this thesis is to address a current gap in research on the use of deixis in the virtual environment: the usage of deictic expressions and gestures that content creators of YouTube gaming tutorials use to establish referencing and communication with their audience. Moreover, my thesis aims to provide additional theoretical integration between gesture research on manual gestures in physical environments and the gestures used in virtual environments. Because I focus on deictic gestures, a discussion of deixis and specifically spatial demonstratives is relevant, especially because demonstratives occurred very frequently in my data. In addition, video game tutorials are an instructive form of text, which means that I will have to discuss speech act theory. For these combined reasons, my literature review is focused on YouTube, video games, gestures, deixis, and speech acts.

This chapter is broken down into seven sections. The first section, 2.1, presents an overview of the video sharing platform, YouTube, and the gaming environment. Section 2.2 provides a rationale for the terminology used in this thesis. Section 2.3 links gestures to pragmatics and is followed by 2.4 which examines and explores the realm of deixis and the concept of the origo. Section 2.5 branches from section 2.4 by exploring the relationship between deixis and speech acts. Section 2.6 builds on 2.5 by discussing spatial deixis and demonstratives. Lastly, section 2.7 discusses Clark's (2003) dichotomy directing-to and placing-for, which have yet to be fully explored in gesture research.

2.1 Computer Games and YouTube

Computer Games

Computer games are a popular mainstream activity, and they make up a large market comparable to that of the movie and music industry (Chik, 2013; Williamson, 2009). There are many reasons people play video games (Yee, 2006): they have been described as being fun, challenging, engaging, and useful for learning (Connolly, Boyle, MacArthur, Hainey & Boyle, 2012), and they provide a place for socializing with others (Lazzaro, 2004; Seay, Jerome, Lee & Kraut, 2004; Williamson, 2009).

As the digital technology used to develop games has continued to evolve, key areas of inquiry have opened up into areas such as the educational implications of games (Cooke-Plagwitz, 2008; Cornillie, Thorne & Desmet, 2012; Gee, 2007; McGonigal, 2012; Ryu, 2013), social gaming (Piirainen-Marsh & Tainio, 2014), gamification (Kapp, 2012), social narratives (MacCallum-Stewart, 2014), language use in gaming (Ensslin & Finnegan, 2019), mediated communication (Wadley, Carter & Gibbs, 2015), the phraseology and lexico-grammatical patterns in video game tutorials and walkthroughs (Gledhill, 2019), the use of avatars in video games to perform deictic gestures (Antonijevic, 2008), and the ways individuals use digital artifacts to indicate (Cooperrider, 2016; Jorge, van Leeuwen, Dams & Bouwen, 2013). However, there are still many areas in the virtual environment, such as YouTube gaming tutorials, that have been explored very little.

YouTube

One way players enrich their experience with gaming is by accessing outside support systems, such as online communities or resources made by other players (MacCallum-Stewart, 2014; Ryu, 2013). YouTube, a form of video-mediated communication, was established in 2005 by Chad Hurley, Steve Chen, and Jawed Karim. Since then, YouTube has become the largest video sharing website on the internet (Duo, 2020; Poche, Jha, Williams, Staten, Vesper & Mahmoud, 2017). YouTube has over 2 billion users which “amounts to almost one-third of the users on the internet” (YouTube, n.d), “with over 50 million content creators” (Ang, 2020), where “89% of YouTube views are from outside the US” (Aslam, 2020). The content posted on YouTube is diverse, ranging from music and videos for entertainment, product and movie reviews, user participation on YouTube (Khan, 2017), as well as videos for educational and academic purposes (Almobarraz, 2018; Burke, Snyder & Rager, 2009; Fralinger & Owens, 2009; Godwin, Khan & Yellowlees 2017; Jones & Cuthrell, 2011; Orus, Barles, Belanche, Casalo, Fraj & Gurrea, 2016; Pecay, 2017). Much of the research on YouTube has focused on self-directed learning (Lee, Osop, Hoe-Lian Goh & Kelni, 2017), gaming (Potts, 2015), vlogging, make-up tutorials (Bhatia, 2018; Choi & Behm-Morawitz, 2017), and advertisements (Dehghani, Niaki, Ramezani & Sali, 2016).

One type of resource individuals post on YouTube is that of tutorials. These are videos for individuals who seek information (Damude, Hoekstra-Weebers, van Leeuwen & Hoekstra, 2017; Khan, 2017; Shoufan, 2018) that may be inaccessible via face-to-face means. These videos include do-it-yourself (DIY) tutorials (Tolson, 2010) that range from installing household appliances to carpet installation and plumbing repairs.

YouTube tutorials are often informal, the creator of the video is more often than not an ordinary person (Choi & Behm-Morawitz, 2017; Tolson, 2010) sharing their knowledge about a specific topic. YouTube content creators also often provide information in order to inspire their viewers to take certain actions by using themselves as examples. By demonstrating how to do something, the viewers can follow along and mimic what the creator is doing. This is the pattern that gaming tutorials follow, videos made by gamers who use themselves as examples in order to share content or information about a game they are playing.

Computer Games and YouTube tutorials

Gaming tutorials fill an important role in the gaming community because they offer advice, guidance, and information about gameplay, especially since the "game designers often provide very little instruction on how to solve in-game problems, providing players with a nearly blank palette from which to explore a huge range of possible solutions based on past experience and intuitions" (Granic, Lobel & Engels, 2014, pp. 69). YouTube gaming tutorials cover a wide range of topics related to games. In some tutorials, content creators explain game mechanics, what is the best way to complete a particular goal, strategies, or simply "to learn how to get better at a game" (Petrova & Gross, 2017). The researchers who focus on YouTube gaming videos mainly investigate spoken discourse, such as the use of sexual innuendo, sarcasm (Potts, 2015), profanity, humor, irony, emotional expression, and casual language in discourse (Lozano-Blasco, Latorre-Martinez & Cortes-Pascual, 2021). Although communication between YouTube gaming content creators and their audience has

been explored in terms of spoken discourse, YouTube gaming videos and tutorials have received little attention on the use of deictic expressions co-occurring with deictic gestures by content creators as a method of communication.

2.2 The Virtual Environment and The Physical World

One of the key issues in current literature on virtual spaces is a debate about the nature of how to refer to these different spaces, with some researchers supporting the terminology of virtual world, while others support virtual environment, and still others support virtual reality. The word virtual is often used to describe a place that is not the physical world but is rather a place created by individuals using technology. Virtual, as Bartle (2003) states, is “that which isn’t, having the form or effect of that which is” (p. 1), where real is “that which is” and imaginary being “that which isn’t” (p. 1). Using these definitions, the word virtual is applicable to computer games, online communities, online chat platforms (Ensslin, 2012), blogs, websites, and video sharing sites like YouTube as well as the content that is uploaded to them. Virtual is a place that is not tangible, instead, it is an environment that is “computed by numbers, as opposed to created by physical matter” (Luciani, 2007, p. 299).

Schroeder (2008; see also 1996) describes virtual environment and virtual reality as: “a computer-generated display that allows or compels the user (or users) to have a sense of being present in an environment other than the one they are actually in, and to interact with that environment” (p. 2). He conflates the definition of virtual environments and virtual reality under the same definition; however, technology has evolved to the point where virtual environments and virtual reality are not the same. Virtual reality,

which is outside of the purview of this research, has its own distinct position that sets it apart from other virtual environments. Taking into consideration the work of the researchers mentioned above, in this thesis I use the wording *virtual environment* as an overarching term to encompass all of the different types of virtual spaces: video games, non-gaming programs, online communities, websites, and virtual worlds. As technology continues to advance, it becomes more and more difficult to define what sets apart that which is virtual and that which is real (see Craig, Sherman & Will, 2009; Golub, 2010; Luciani, 2007; Mallon & Lynch, 2014; Whitton, 2003), for this reason, whenever I refer to the world that people physically live in, I will use the term *physical world* rather than real world.

2.3 Gestures

Research on gestures covers a wide range of topics and genres, such as pointing with children of different ages (Butterworth, 2003; Goldin-Meadow & Butcher, 2003; Masataka, 2003), co-occurrence of gesture and speech (Kendon, 2004; Kita, 2003; McNeill, 1992), bodily displays used to make referencing salient (Eriksson, 2009), pointing as situated practice (Goodwin, 2013), deixis (Abdullah, 2015; Bühler, 1990 [1934]; Lenz, 2003; Lyons, 1977), gestures in face-to-face interactions (Holler, Kendrick & Levinson, 2017), gestures in L2 acquisition (Gullberg, 2006; Kellerman, 1992; Lee, Hampel & Kukulska-Hulme, 2019), and virtually embodied actions, such as avatar use in manipulating and exploring a game (Newon, 2011). Out of this research has come a commonly accepted definition of a gesture as using some part of the body to communicate or indicate so as to be understood by a recipient (Subramani, 2010).

Gestures are “movements of the hands and arms that we see when people talk,” they are “spontaneous creations of individual speakers, unique and personal” (McNeill, 1992, p. 1). They are understood to be visible bodily action that is seen by interactants to contribute to the ongoing discourse (Kendon, 2004) and at the same time, “…when speech is unavailable…gesture can become a form of language all by itself” (ibid, p. 3). From these it can be generally understood that gestures are bodily movements that are part of linguistic utterances, where utterances are “any ensemble of action that counts…[as] an attempt by the actor to ‘give’ information of some sort” (Kendon, 2004, p. 7). This distinguishes gestures from body movements that are not communicative. Scratching one’s arm during a conversation, for example, while being a movement that may co-occur with speech, is not commonly interpreted as being communicative. It is simply a body movement.

Kendon’s (1997, 2004) seminal work focused on how visible bodily actions, or gestures, are used in utterance construction as part of the spoken discourse. Prior research on gesture thought of it as an ‘add on’ to speech, however, Kendon was one of the first to link gesture and speech to each other. He showed how they are “two aspects of the process of utterance” (Kendon, 2011, p. 207). In addition to this, Kendon provided a breakdown of the different movement phases of a gesture. He postulated the existence of three different phases that occur in gestures that take place in face-to-face contexts. These phases are the *preparation* phase, the *stroke* phase, and the *recovery* phase and deal with the positioning of the hand that is making the movement. The stroke phase consists of the main part of the actual gesture and is generally the point of highest velocity, where the most movement occurs, and is generally semantically

connected to the concomitant speech context. In contrast, preparation and retraction phases are not intrinsically meaningful themselves, but exist merely to facilitate the execution of a stroke (preparation) and resume hand position after it (retraction).

Kendon's (1988) work also influenced McNeill to examine gestures in light of the degree to which they are used with spoken language. McNeill developed what he called 'Kendon's continuum'. On the far left side of this continuum is gesticulation and on the far right side is sign language. As one moves from gesticulation to sign language on the continuum, the "obligatory presence of speech declines, the presence of language properties increases, and idiosyncratic gestures are replaced by socially regulated signs" (McNeill, 1992, p. 37). McNeill developed the continuum to showcase how gestures are not all the same but rather are "distinguish[able] among different categories" (McNeill, 1992, p. 37; see McNeill, 1992, chapter 2 for details).

McNeill (1992) also introduced a widely cited taxonomy of gesture (see Kendon, 2004, chapter 6 for a history on gesture classification). He described five main types of gesture movements: iconic, metaphoric, beats, cohesives, and deictic. Iconic gestures closely resemble the content of the speech that is produced. They are used to illustrate concrete objects or actions. For example, if a speaker says, "go up the stairs" and performs an ascending walking gesture with their index and middle finger, this gesture represents a person walking up stairs. Metaphoric gestures are similar to iconic, but instead refer to abstract objects or actions. For example, "someone can complain about an excessively long lecture, not only by saying the lecture goes on and on, but also by adding a gesture, rolling his/her hand at the same time" (He, Nagels, Schlesewsky & Straube, 2018, p. 1). Beats are "simple, rhythmic gestures that do not convey semantic

content” (Alibali, Heath & Myers, 2001, p. 169) and instead “place emphasis on the particular word or phrase they accompany with a baton movement which is often metronomic” (Bernard, Millman & Mittal, 2015, p. 4017). Cohesives connect related but temporally separated parts of discourse (McNeill, 1992, p. 16). For example, to refer back to a previously mentioned topic, the speaker may wave their hands in the direction behind their shoulders. Deictic gestures are used to indicate referents, and often take the shape of a pointed finger but any “extensible object or body can be used” (McNeill, 1992, p. 80), such as a laser or telescopic pointer. From McNeill’s gesture types, I have chosen deictics as the focus of this study and will explore them in more detail in section 2.4.

2.3.1 Gestures and Pragmatics

Kendon and McNeill’s typologies, theories, frameworks, and methods have been used in researching how gestures affect face-to-face discourse and pragmatics, as seen in the work of Lewandowski and Ozcaliskan (2018), Lin (2017), and Wessel-Tolvig and Paggio (2016). Pragmatics is “the study of how more gets communicated than is said” (Yule, 1996, p. 3). Pragmatics “studies the way in which the context shapes meaning…the interpretation of utterances not only depends on the linguistic knowledge, but also…knowledge about the context of the utterance” (Stapleton, 2017, p. 2). Gestures play an important role in pragmatics because they are one of the devices that add more to the communication than what is simply uttered, and they “can be used to provide context for spoken expression, thus reducing the ambiguity of the meaning of what is expressed” (Kendon, 2000, p. 60). Gestures are one of the ways in which

speakers can enrich the semantic content of utterances and they are conditioned by pragmatic context (MacLeod, 2009).

In the field of pragmatics, gestures have been shown to be a way to repair communication breakdown. For example, iconic and metaphoric gestures can occur in cases where the topic is unknown or difficult for the participants by “standing in for concepts they [the participants] did not yet articulate in speech” (Singer, Radinsky & Goldman, 2008, p. 377). Gestures can help with word recall (Goodwin & Goodwin, 1986) and they can help a speaker “formulat[e] what to say” (Matsumoto & Canagarajah, 2020, p. 245). Gestures also “exhibit images that cannot always be expressed in speech, as well as images the speaker thinks are concealed” (McNeill, 1992, p. 11). Similarly, “gesture can convey information that is difficult to formulate verbally” (Feyereisen, 2017, p. 58).

Gestures are used in pragmatic discourse to reduce the ambiguity of referents (Cooperrider, 2016). An example of this is when a speaker makes a request for an item by directly pointing to the object being requested so that it is clear to the interlocutor which object they want. Gestures can also be used as a *chain of indication* (Clark 2003) or *nonliteral pointing* (Morford & Goldin-Meadow, 1997) both refer to an object without it being the focus of the gesture, but rather the association that object has to another person, place, or object. For example, while tapping on a *Harry Potter* book the speaker may say, “Did you know she wrote other books as well?” where “she” refers to the author, who is not presently there. The pointing is referring to who or what the book represents, in this case the author.

A further pragmatic function of gestures is to “express the attitudes of the speakers to their utterances or to their addressees” (Feyereisen, 2017, p. 53) and gestures can play an important part in conveying information, such as when gestures are used to complete sentences in which words are omitted. An example can be seen in the following sentence from McNeill (2006): “Sylvester went [gesture of an object flying out laterally]” (p.2, *speech-framed gestures*). In this example, without verbally expressing or describing it, the speaker shows with gestures the direction and manner of Sylvester’s motion. Finally, it is important to keep in mind that gestures are a “medium of expression that humans have at their disposal which they can use for a wide range of different expressive purposes. What forms of gesture are created and used depends upon the circumstances of use, the person’s specific communicative purposes and what other modes of expression are also available” (Kendon, 2004, p. 84).

In addition to pragmatics, research on gesture has been examined in a variety of other fields, the most important for this research is Human-Computer Interaction (HCI) (Lee, Jeong, Lee & Park, 2009; Nickel & Stiefelhagen, 2003; Oleksik, Milic-Frayling & Jones, 2014). In pragmatics, gestures are studied in order to understand how they create meaning; however, HCI research does not necessarily work with a linguistic concept of gesture. In linguistic research, gesture is understood to be something that is communicative (i.e., non-communicative movements are not gestures). Body movements in HCI applications are functional and goal-oriented movements that are needed to interact with the computer system; they are not necessarily communicative in the sense of being used as part of conversation. Exploring HCI, however, is still relevant

because it is a subsection of gestures research that shows how there can be a blurring of what is functional and what is communicative.

Researchers in HCI have examined how gestures can be mapped onto recognition systems that can then be implemented in different technological devices, programs, or contexts, such as “electronic entertainment, video surveillance, patient monitoring, nursing homes, [and] smart homes” (Zhao, Li, Pang, Sheng, Wang & Ye, 2014). In the gaming industry, these gesture recognition systems are necessary as many gaming devices use the systems to collect the data of human movement and then code it into the games and hardware. One example is the Xbox Kinect, which is a motion sensor for the Xbox game console. A controller is not needed to play Kinect games as the human body itself is the controller. When the hand moves, so does the cursor on the TV screen and this enables players to interact with the console in various ways, such as changing settings or playing games.

In some areas of HCI, there are researchers who believe the mouse and keyboard are “almost totally free of gestures” (Aimaiti & Yan, 2011, p.1), that they are mere functions, tools that are manipulated by users to complete tasks on the computer. However, they took hand movements that are functional, such as those used with the Kinect, and considered them to be gestures (Cai, Zhu, Wu, Liu & Hu, 2018; Zhao et al., 2014). It is possible that some have this belief because traditionally the mouse and keyboard are input devices used functionally to move things forward or to get things done. However, when used with devices that have gesture recognition programs, bodily gestures are also simply input devices: functional tools to move things forward or get things done. How then, do we determine when gestures are communicative and when

they are not? This question was answered by the work of Novack, Wakefield and Goldin-Meadow (2016) who focused on how humans interpret the actions of others. In this work, the researchers describe three types of hand movements. The first is extrinsic goal-oriented movement, functional movements that an individual uses to manipulate the surrounding world, such as picking up and moving an object from one place to another. The second is “movement for the sake of movement” (p. 340), in which the movements are just movements and are neither for communicative purposes nor for manipulating the surrounding world. An example of this would be dancing or exercising (Novack & Goldin-Meadow, 2017). The third type is gestures, which are movements of the hand that convey meaning.

The differentiation proposed by Novack et al. (2016) between the types of actions of a hand movement can be extended to the virtual environment as well. Using the Xbox Kinect as an example, when a person gestures with their arm and hand to open the settings or other options on the console, the players are goal-oriented to do something. They are not adding to the discourse or providing meaning to an utterance. This is analogous to using a mouse to open settings and other options on a computer. The only difference in this case is that instead of the traditional mouse arrow appearing on the screen of the TV for the Xbox console, a hand icon appears which follows the player’s bodily movement (Figure 2.1).

Figure 2.1 YouTube handle: Xbox on – Title: Getting Started with Kinect – The Basics [PEGI 3] URL: https://www.youtube.com/watch?v=ELrEJJT_eng

Even though HCI research focuses on a different notion of ‘gesture’, it is relevant to this research because if a bodily movement used by an individual for functional purposes on a gaming console, such as the Xbox, is considered a gesture (Kim, Kim & Yoo, 2013; Aimaiti & Yan, 2011), then a mouse pointer has the potential to be considered a gesture as well, as they not only move things forward, but are also communicative devices in the virtual environment. As Spagnolli, Gamberini, and Reno (2005) state, “pointing can be performed with any bodily-connected device producing an index, expanding the definition of pointing to legitimately include the mediated one...Mediated ‘movements’ can be considered as ‘gestures’...” (p. 71). The use of the mouse in virtual environments, such as that of YouTube gaming tutorials, has qualities of both meaning making, as seen in pragmatics, and functional purposes, as seen in HCI.

Mouse movements can provide additional information to what is verbally uttered by directing viewer attention to the particular part of the screen that is being featured in the discourse. By using the mouse pointer as a way to indicate objects on the screen, the speaker can remove ambiguity as to what they are referencing. An example of the dual quality of the mouse can be seen in Due and Toft (2021) who found that the use of

a mouse was a way to establish joint attention amongst interlocutors and to remove ambiguity of what was being verbally uttered. To indicate to the interlocutor what was being referenced, the mouse was used to directly point at and highlight text. It was also used to edit by selecting the text and then deleting it. The mouse being used to select the text in this context could be argued as having a dual use, to indicate and to highlight text to be deleted. The mouse was a functional device to move things forward, e.g., delete text, and also a pragmatic device where the deictic pointing and highlighting were used to establish referencing, e.g., which text to delete, thereby giving the utterance meaning by tying it to the context and thus establishing communication (see also Lee and Tatar, 2012). The mouse in this context is an essential part of the discourse and is used in a meaningful way, and this opens up key questions about how far technological movements and affordances can be considered as gestures, and if so, what kinds of functions they perform.

2.4 Deictics

Deixis is one of the primary methods people anchor communication to context, and it is a “way entities and activities are referred to in terms of the shared context of interlocutors” (Zappavigna, 2019, p.5). Deixis creates meaning that can be understood by all interlocutors involved in the discourse and it tells us “where to look for the particular item that is referred to” (Mey, 2007, p. 54). Speech that is used to verbally indicate and reference is known as a deictic expression (Hanks, 2017; Yule, 1996), and contains terms such as “this”, “that”, “he”, “she”. When deictic expressions are used, “the speaker refers to a certain entity” or object (Lenz, 2003, p. vii) and context is

required for the expression to have meaning. This context is provided by the deictic field (Bühler, 1990).

The deictic field is separated into three main types of axes: personal, temporal, and spatial (Figure 2.2). The personal axis generally involves two types of deictic expressions, either personal pronouns such as “I, you, he”, or honorifics such as “Miss, Ms. and Mr.” (Levinson, 2006, p. 119). The spatial axis refers to location and involves deictic expressions such as “here and there” (Ensslin, 2012, p. 101). The temporal axis refers to the flow of time, involving such deictic expressions as ‘now’ and ‘later’ or ‘yesterday’ and ‘today’. This deictic field is where the participants, objects, events, locations, and times are all tied together to create meaning. The field generally emanates from the location of the speaker, who is usually the reference point, or the *origo* (Bühler, 1990).

The origo is the center of an utterance, everything that is referenced is indicated from that point. For example, if I say, “I want that cup” and pointed to it, the point and the use of the deictic expression ‘that’ is being used and indicated from *my* location, *my* perspective, in relation to where *I* am. If a speaker is referencing themselves in relation to a physically close object, they may use personal, spatial, and temporal terms such as ‘I (personal) am (temporal) here (spatial)’. If the object of reference is far away, the speaker may use speech that relates to the far distance, such as ‘she (personal) is (temporal) over there (spatial)’. One caveat to Bühler’s concept of origo is that it is mainly based on the people and objects found in the physical world. However, the concept of origo has changed with the presence of the virtual environment, and this will be further discussed at the end of this section.

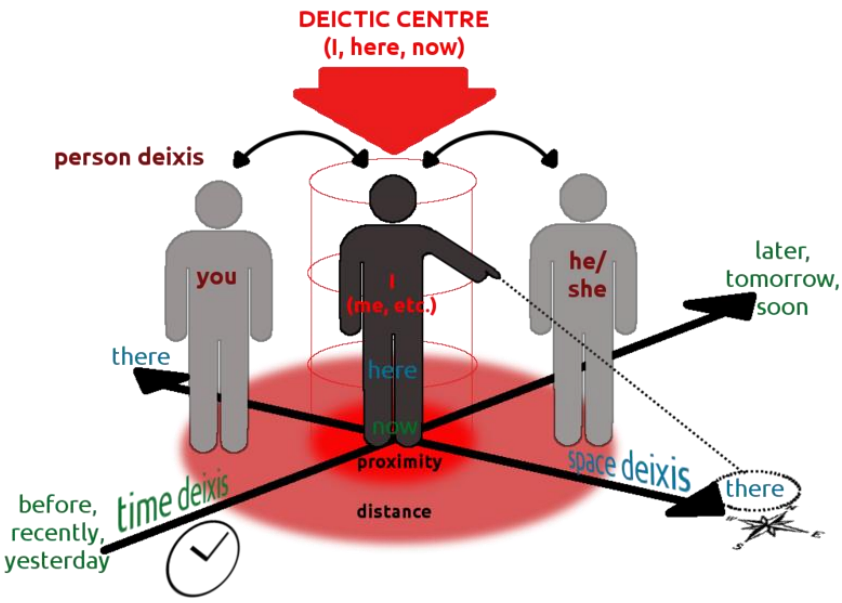

Figure 2.2: Deictic field (Wesn, 2013).

Deictic gestures are another way deixis is performed, and they can complete or disambiguate the intention of deictic expressions. Deictic gestures are movements of the body, such as the index finger or a laser pointer, that visually indicate a referent. Deictic gestures in the physical world are a "highly conventionalized and well-established way to solve the problem of identifying the referent…" (Eriksson, 2009, p. 252), especially since they are "...used to indicate person[s], objects, directions, or locations…" (Krauss, Chen & Gottesman, 2000, p. 262-263). They are body movements that form part of linguistic utterances (Kendon, 2004), and are necessary for referencing or indicating an entity "in relation to the spatiotemporal context created and sustained by the act of utterance and the participation in it..." (Lyons, 1977, p. 637).

Deictic gestures are a fruitful area of research because pointing gestures, a mainstay of deixis, are "a foundational building block of human communication." (Kita,

2003, p. 1). In fact, Kita (2003) goes on to list four additional reasons for why studying deictic gestures is important:

> First, it is ubiquitous in our day-to-day interaction with others. When communicating about referents locatable in the speech situation, pointing is almost inevitable…Second, pointing is a uniquely human behavior. In other words, pointing separates humans from primates, just like the use of language does…Third, pointing is primordial in ontogeny. Pointing is one of the first versatile communicative devices that an infant acquires… Fourth, pointing does not merely indicate a vector, but it can serve to create further types of signs. For example, a pointing gesture can create an iconic representation by tracing a shape or movement trajectory… (Kita, 2003, pp. 1-2).

Deictic gestures are “clear precursors to important linguistic and communicative functions, including establishing reference and promoting shared attention” (Frank, Braginsky, Yurovsky & Marchman, 2021, p. 2). There is always the possibility of “miscommunication and misperceptions based upon differing norms of interaction across societies and speech communities” (Boxer, 2002, p. 150), however, deictic gestures can be used to alleviate these issues of communication. If an individual would like an object, they only need to point to it, even if they do not know the word for it in the other language, the recipient of the deictic gesture would be able to understand what the person is referencing. Deictic gestures have an important role in discourse as they are not only used to indicate, but also contribute to the conversation. Lin (2017) examined the relationship between gesture and speech by participants from three

different countries (Taiwan, India, Indonesia), and found that their participants would often use deictic expressions and gestures during discourse to direct attention to objects, to talk about location, direction, and time, or to refer back to a previously mentioned person, object, or event (see also Gullberg, 2006, and Stam & McCafferty, 2008).

In most cases, both deictic expressions and deictic gestures are used together in an utterance because they rely on each other to unambiguously indicate a reference (Cooperrider, 2016). They are also an important component in social interactions because they are used to organize, establish, and coordinate joint attention, in which the speaker and the addressee both attend to the same object of reference (Diessel, 2006; see also Lee, Hampel & Kukulska-Hulme, 2019). In addition, they are used to guide "the attention of the addressee…towards the referent" (Lenz, 2003, p. viii). In other words, deictic expressions and gestures are crucial in pragmatics because they can be used to decipher pragmatic contexts and discourse, by linking together people, places, and objects.

Deictic gestures and expressions can also be used to reference objects that are not within the interlocutors' physical surroundings, objects that are not visible (Stukenbrock, 2014), objects that are not tangible, such as memories and conceptualized scenes (see Bühler, 1934/2011), and objects that are "spatially removed but already known to both parties" (Hanks, 2009, p. 22). Furthermore, they can be used in negotiation, conflict resolution (Glover, 2000), and to bridge communication gaps (Gullberg, 2006).

Without deictic gestures and expressions, communication and referencing could be extremely difficult, especially “when problems occur in language comprehension as well as in language production” (Klippi, 2015, p. 337). If someone wanted to reference an object without the use of any deixis, they would need to find an alternative way to indicate it. Descriptions about the object’s features or location could be used, but it takes time to describe items. In addition, having multiple potential referents in the same vicinity as the target reference could make it difficult for the addressee to find it, regardless of the object’s proximity to the speaker (Hindmarsh & Heath, 2000). Using only descriptions to indicate places a cognitive demand on the addressee, who must determine which of the potential objects of reference in a general direction fits the criteria of the description (Piwek, Jan Beun & Cremers, 2007). For example, if a speaker and addressee are in the same room, and the speaker says, “hand me that small red box”. The addressee must look around for an object that is small, red, and box shaped. However, if the speaker just points to the object, the addressee can look in that direction and find the object. The addressee does not need to think which object is the referent because it is being directly pointed to. Using a deictic gesture to indicate makes it much easier and faster to reference, and in some cases, it can make what the intended target is even clearer, especially in conjunction with deictic expressions and other bodily movements and placement (see Hindmarsh & Heath, 2000).

Origo

As stated previously, origo is an important factor when determining the location of a referenced object, as referencing is made in relation to that center. Although

referencing from the origo is generally from the perspective of the speaker, Bühler (1990 [1934]) states that the origo can shift, or move, to other people, creatures, and objects. This means that when a reference is being made, the personal, spatial, and temporal language used can be from the position of another entity that is not the speaker. In other words, the one speaking does not necessarily have to be referencing things from their position, but rather they can make a reference from another entity's origo, or position. The below paragraph is an example of this taken from Bühler (1990 [1934]):

> If I am standing as the commander nose to nose in front of gymnasts lined up in a row, convention requires that I choose the commands 'forwards, backwards, right turn, left turn' not according to my own orientation system, but to that of the others, and the translation is psychologically so simple that every group leader learns to master it. (Bühler, 1990, p.118).

This change in origo positioning is something that can also be seen in virtual environments. Take, for example, Figure 2.3, which shows a screenshot of the game *The Elder Scrolls Online*. In Figure 2.3 A, the avatar has their back facing the game player, whereas in Figure 2.3 B, the avatar has their front facing the game player. If the game player were to refer to the container from the perspective of the avatar, the container would be on the avatar's right side, as seen in Figure 2.3 A. However, in Figure 2.3 B, the container would be on the avatar's left side. This is the case because we are using the avatar's position to reference the container, not our positioning as the game player looking at the computer screen. On the other hand, from the perspective of

the game player and audience who are looking at the screen, if they refer to the container, they would refer to the chest as being on the right side regardless of avatar positioning in either image. The avatar's positioning in this case does not matter for the viewers or players in regard to referencing the container, since the origo is from their perspective not from the perspective of the avatar.

Figure 2.3 The Elder Scrolls Online: Left picture (A): The avatars' back is facing the player. Right picture (B): The avatars' front is facing the player.

Bühler's origo shift has also found its way into more recent studies that focus on the virtual environment. This shift in origo perspective is similar to the phenomenon of player immersion within digital fiction explored by Bell, Ensslin, Van Der Bom, and Smith (2018). Digital fiction is a narrative that is accessed through a digital medium, often similar to a 3D video game. The authors found that under certain conditions, players feel as if they are in the game itself, even though physically they are not. The authors explain this phenomenon using the Deictic Shift Theory (DST – see Duchan,

Bruder & Hewitt, 1995), which “demonstrate[s] how readers are completely engaged in narratives, to a degree that they interpret events in narrative as if they were experiencing them from a position within the story world” (AbdulilahGheni, 2019). Bell et al. (2018) demonstrated that DST occurred when the players in the physical world embody their avatar in the digital and “shift… imaginatively into … [the] ontological domain” (p. 6) of the digital story world.

Bell et al. (2018) also incorporate Ensslin’s (2009) *doubly-situated* (see also Herman, 2002) theory which refers to a player referencing themselves simultaneously both as their physical self and as their avatar in the game. In these cases, there is a spatial deictic shift. The way the Deictic Shift Theory and doubly-situated are used illustrate their similarity to Bühler’s shift in origo. The difference between the three is that in DST a player is located in the physical world but is immersed, or shifted, into another domain. They interpret and experience it as if they were there. Doubly-situated is used to reference both oneself and the avatar in the other domain simultaneously, thus the player is interpreting themselves in both domains. In contrast, Bühler’s origo shift originally did not refer to individuals going between worlds. It was originally formulated as a simple change in perspective, the deictic expressions being spoken from the speaker’s point of view or from another person or objects. It seems that the origo shift of Bühler is a foundation for the other two theories, and that the phenomenon it explains has merely adapted to new environments as new technology has developed.

When it comes to gaming videos and tutorials, the concepts of DST, doubly-situated, and the shifting of the origo are all present. For example, if the pronoun ‘you’ is used by a player addressing another player within a game, the pronoun could be

referring to either the avatar, the person controlling the avatar, or both (Abdullah, 2014). Previous research on these concepts have explored an individual's immersive experience in a game, such as those made specifically for narrative story telling (Colthup, 2018), a player deictically shifting into the first-person perspective of the avatar (Won, 2019), or using personal deixis in certain mediated genres (Bell et al., 2018). However, questions remain about how these concepts might relate to the mediated gestures that are part and parcel of YouTube gaming tutorials. As such, this leaves an opening for further research on how they can be applied to this medium, and how spatial deictic expressions and deictic gestures play a part in tying deictic shifting to the context.

2.5 Speech Acts and the Co-occurrence of Deictic Behaviors

A key concern in gesture research is their communicative function when combined with verbal content. Given that the data in this thesis are YouTube gaming tutorials, the way language is used to help players achieve particular goals is crucial. One way to approach this is to consider how gestures co-occur with speech acts. Speech acts are "the performance of a certain act through words" (Gass, 1996, p.1), where the "uttering of a sentence is… an action within the framework of social institutions and conventions" (Huang, 2014, p. 119). In essence, a speech act is when people use words to perform or cause action. I chose to use Searle's (1976) classification of speech acts, rather than Austin's (1962), which divides them into assertives, directives, commissives, declaratives, and expressives. Although Austin was a foundational source for Searle's classification, Searle is largely considered as the starting point for analysis of speech

acts (Huang, 2014) as Austin's categories have been criticized for not being sufficiently distinct nor clearly separate from each other (Mey, 2007).

Although speech acts have been studied in relation to gestures in face-to-face discourse (Ciroux, 2020; Miyake & Sugimura, 2017), most research that investigates speech acts and deixis still focus heavily on the linguistic aspect (Liang, 2021; Yulfi, 2017). For example, in politeness theory, Haverkate (1992), Koike (1989), and Trommer (2011) investigated personal and temporal deixis and how it is used to create degrees of politeness in speech acts. Although they used different approaches, their results were the same: deictic expressions that express distance from a speaker's origo are more polite than those used for close proximity. The commonality between Haverkate (1992), Koike (1989), and Trommer (2011) is that they demonstrate a link between deixis and speech acts in face-to-face settings. However, the researchers did not discuss gestures in relation to speech acts, or the relationship between spatial deixis and speech acts. The only reference to this interplay was where Haverkate (1992) states that "place deixis…does not seem to play a specific part in the expression of mitigating strategies" (p. 508).

On the other hand, speech acts and their application in the virtual environment have emerged in recent studies, such as with live YouTube gaming stream chats (Kurniasih, Falaakh & Setyaningsih, 2022) and between interlocutors interacting in an online game (Liang, 2021). Taking a qualitative approach, Kurniasih et al. (2022) found that most individuals who commented in the chat used expressive speech acts. Similarly, Liang (2021), observed *Second Life* game players who would use avatar positioning and movements to mitigate the face threat of directive utterances. Kurniasih

et al. (2022) and Liang (2021) both examined gaming and its relationship with speech acts, and their findings show that there is interplay between deixis and the virtual environment, however neither discussed how the speech acts used by their participants were tied to their context, or if deictic gestures played a role in their data. This opens potential for further inquiry into how speech acts tie to their context in a gaming environment, and furthermore, how deictic expressions and gestures can be used to establish this connection within YouTube gaming tutorials.

In addition to the work discussed above, Alibali, Young, Crooks, Yeo, Wolfgram, Ledesma, Nathan, Church, and Knuth (2013), Bamford (2004), Bell et al. (2018), and Nordvall (2012) all discuss deixis in different contexts, but speech acts were not examined for any effect they may have had even when they occurred in the data. Deictic expressions and gestures are important aspects to consider when examining speech acts, especially in a gaming context, as deictic expressions are used in conjunction with speech acts “to achieve something, to address or to discriminate, or simply to distinguish” (Widlok, 2015, p. 90) while deictic gestures are used to visually point out where the object of reference is located. Both deictic gestures and expressions are used in speech acts to ground the utterance to the context, as well as to direct the interlocutor’s attention to the referent. To my knowledge, the relationship between speech acts, deictic expressions and gestures to establish communication in YouTube gaming tutorials has yet to be fully explored. The work in this thesis attempts to address these gaps and go beyond the current literature by incorporating these facets into the analysis.

2.6 Spatial Deixis

The following section focuses on exploring the foundation of spatial deixis and understanding the deictic expressions that are found within this category. I have chosen to focus on spatial deictic expressions because spatial deixis is used to “identify the direction of the motion, towards or away from the place of [the] speech event…spatial deixis is the marking in language of the orientation or position in… space” (Stapleton, 2017, p. 5) and this can be seen both in face-to-face gaming as well as YouTube gaming tutorials. An example of this in a game setting has been explored by Nordvall (2012). The players in his research used spatial deictic expressions and gestures to communicate and indicate objects to each other when they were playing the board game WarCraft. Using both deictic expressions and gestures provides more clarity to the other players because the deictic expression not only grabs the attention of the others, but it also directs the other players to the object of reference verbally, while the deictic gesture clarifies the target visually. An example of this was when a player would indicate which of their pawns would receive a resource. In examples such as this, the player would often use directives with deictic expressions, and this combined use with deictic gestures would influence other players who were sharing in the joint activity to perform actions. In the case of the previous example, after the first player indicated the pawn that would receive a resource, both verbally and with a deictic gesture, the other player handed a gold coin to them. His research showed that there is a relationship between spatial deictic expressions and gestures in face-to-face gaming, and this opens opportunities to investigate the relationship between the two in virtual games and by extension online gaming tutorials.

There are two types of verbal referencing in English: deictic expressions, such as demonstrative pronouns “this” or “that”, and nominals, which are nouns or noun phrases, such as “Mary” or “the chest”. Demonstratives are usually used in a way that “instructs the hearer to attend to something beyond the talk itself, that is, the [manual] point, to locate what is being indicated…” (Goodwin, 2003, p. 223). Demonstratives are universal, they are some of the first words children learn, and they are often “accompanied by a deictic pointing gesture” (Diessel, 2006, p. 469). In fact, demonstratives, “serve the same function as a deictic pointing gesture. Both a demonstrative and a deictic pointing gesture indicate the location of an object relative to the deictic centre [origo]…they provide spatial orientation, and both function to focus the addressee’s attention on a particular referent” (ibid, p. 470). Demonstratives are “one of the central ways in which language can be used to coordinate attention and enable social interaction” (Rocca & Wallentin, 2020, p. 1). In contrast, nominals can establish reference in a more self-contained way; they do not necessarily need a gesture because the object itself can usually be distinctly identified from the content of the utterance. Nominals are not deictic expressions in the sense that they do not require context to gain meaning. For example, in the utterance, “the stapler next to the book”, “the stapler” is a nominal, and it is clear from the utterance that this object is the reference.

Demonstratives can be further broken down into two categories: proximal and distal. Proximal refers to deictic expressions that indicate objects that are in some manner close to the speaker while distal refers to deictic expressions that are in some manner distant from the speaker. For example, “here” is a spatial-proximal expression

while “there” is a spatial-distal expression (Ensslin, 2012). These two distinctions, proximal and distal, are usually used depending on the speaker’s point of view. For example, by using “this” in a sentence, an interlocutor may look for a referent that was near the speaker’s vicinity, whereas with the use of “that”, the interlocutor may look for something beyond the speaker’s vicinity (Wu, 2004). The idea that distance plays a role in demonstrative choice is not a new concept, nevertheless, there are other factors that should be considered.

Piwek, Beun, and Cremers (2008) state that the choice of demonstrative is not necessarily based from a linguistic perspective but rather from the “cognitive mechanisms underlying language production and comprehension” (Piwek et al., 2008, p. 694). In their research they hypothesized that the choice of proximal and distal may be dependent on a variety of factors, one of which was how important the object of reference was for the speaker. On the other hand, Glover (2000) hypothesized that proximal deictics were used to reference objects that were either seen as “problematic or unresolved” (p. 918) or simply topics that were new to the discussion. Mesh, Cruz, van de Weijer, Burenhult, and Gullberg (2021) found that the coordination of demonstratives and pointing changed as the scale of the search space grew. Lastly, Rocca and Wallentin (2020) found that participants tended to use the proximal “this” for manipulatable objects that were small, harmless, and inanimate, e.g., apples, whereas participants would use the distal “that” for objects that were not manipulatable, such as large, harmful, and animate objects, e.g., a tiger.

Cooperrider (2016) performed an experimental study that was the first to show that ambiguity of reference was a dimension that was previously hidden in studies of

proximal versus distal demonstrative use. When we point, distance and ambiguity are naturally correlated: objects we point to that are further away are also less clearly identified by our point. To disentangle these two dimensions that are naturally correlated, Cooperrider (2016) asked participants to refer to objects while pointing with a laser pointer, which allows unambiguous points at a distance. The results showed that in the laser pointer condition, people used proximal rather than distal demonstratives, even though the objects were far away. This suggests that the use of proximal deictics may largely be about ambiguity and not necessarily about distance per se, and previous studies that have claimed the proximal and distal distinction to be about distance have not sufficiently decoupled this from ambiguity.

The main findings by Cooperrider (2016), Glover (2000), Mesh et al. (2021), Piwek et al. (2008), and Rocca and Wallentin (2020) was that distance does not necessarily account for demonstrative choice or pointing, but rather the choice may be influenced by other factors such as a prior conditioning, social-pragmatic factors, characteristics of the object, ambiguity or individual differences such as “preferences, experiences, and personality traits…” (Rocca & Wallentin, 2020, p. 9). Though theories explaining for demonstrative choice are present, there has been little to no attention on how these concepts apply to a virtual environment. This is especially important because the argument of demonstrative choice being based on distance does not apply in gaming tutorials because there is no distance between speakers and objects, as everything referred to in the tutorial is contained on a screen and can be pointed to by the mouse. These gaps in literature, as well as the distinctiveness of the screen mediated virtual environment, have led me to examine demonstrative use in gaming

tutorials using frameworks that have already been established, particularly Cooperrider's (2016) framework on ambiguity (which will be discussed further in Chapter 5).

2.7 Clark's Directing-to and Placing-for

As will be made clear in the results section, I will rely heavily on Clark's (2003) theoretical work on two different types of indication: directing-to and placing-for. Clark (2003) was the first to make the distinction that attention factors into deixis in two fundamentally different ways, either by changing attention with respect to the world, or by changing the world with respect to people's attention. No other work in gesture focuses on these two different ways of orienting with respect to people's attention. This distinction manifests itself in a number of different deictic behaviors, such as when a speaker points to something with their index finger (directing-to) or nudges a coffee cup forward in a diner to request a refill (placing-for).

Directing-to and placing-for are both used to indicate, by "speakers try[ing] to establish a connection, often spatial, between the index they create and the object of that index" (Clark, 2003, p. 247), but they fundamentally differ in the role of the interlocutor's attention. Directing-to directs attention whereas placing-for places an object into an existing field of attention. Directing-to occurs when "people [use] an action as a directive to their addressees to focus their attention on an object, place or event," (Clark, 2005, p. 509). In other words, directing-to is when the speaker directs the participant's attention to the object of reference.

Directing-to is rarely used without language (Clark, 2003; Gerhardt, 2019). Instead, it is often used with “composite signals” (Clark, 2003, p. 253) such as demonstrative pronouns and adjectives. For example, if a speaker pointed to an object, without any words, the addressee might not know why the speaker is pointing at the object, or what to do with it, aside from maybe handing it to them if it is tangible. Similarly, if a speaker were to only use a deictic expression without any deictic gesture, the listener may not know what exactly it is the speaker is referring to. Consider the following example: “Duncan: [Pointing at a painting by Picasso] That’s beautiful” (Clark, 2003, p. 253). The direct point in this example anchors the deictic expression to the context.

An archetype of directing-to is usually a pointing gesture, which is often accomplished with the index finger (Krauss et al., 2000) is a simple and direct way “to indicate person, objects, directions, or locations…” (Clark, 2005, pp. 262-263). Pointing is usually relatively unambiguous and salient, where someone “sticks [their] finger out in the appropriate direction, perhaps saying some accompanying words, and [the] interlocutors follow the trajectory of [the] arrow-like digit to the intended referent” (Haviland, 2000, p. 14). Although deictic gestures that are a type of directing-to usually take on the form of a direct point, there are also other ways to indicate directly, such as tapping with the finger on an object of reference, use of other body parts, conspicuously diverting one’s eye gaze to an object to draw attention to it (Clark, 2003), or as seen in some cultures, the use of the lips to point, rather than the fingers (Cooperrider, Slotta & Nunez, 2018). This attests to the fact that different body parts can be used to perform

the act of directing other people's attention for communicative purposes, something that can be seen to extend into the virtual environment.

Placing-for occurs when "people place an object or action for their addressees at a specialized site as a directive to interpret the object or action appropriate to that site" (Clark, 2005, p. 510). In a placing-for, the speaker brings the object of reference into the view of the interlocutor. Taking an example from Clark (2003), when a customer places objects on the checkout counter, the cashier who sees this knows that the customer would like to buy the items. The object is placed into the view of the cashier, which draws their attention to the object. Both the customer and cashier understand that the place where they are located and interacting, the checkout counter, is a place where certain actions are to take place: the customer places items they want to buy, the cashier rings up the items, the customer hands money to the cashier, the cashier places the items into a bag and hands it to the customer.

The location where the interaction between the clerk and customer is taking place is referred to as an "indexing site" (Clark, 2003, p. 249). An indexing site is a location where the participants in an interaction reference object(s) and take actions that are appropriate for that site, e.g., a checkout counter being a site where transactions take place. By placing the objects on a site where specific actions take place, as well as the shared common ground that both the cashier and the customers have, the interactants understand what type of exchange will happen. Verbal deictic expression is not necessary for this exchange; however, it does require the participants to be aware of what is happening. If either the cashier or the buyer is not paying attention, it may be necessary for either party to get the other's attention and in these cases, deictic speech

may be necessary, for example, Cashier: “are you buying these,” or buyer: “I’d like to purchase these two things over here” (see Clark, 1996).

Besides being different in how they act on the attention of the addressee, Clark contends that directing-to and placing-for differ along several other dimensions. They each have their own roles and are used in specific contexts for specific purposes. The first difference is with respect to how they orient the addressee’s attention, as discussed above. The second difference between the two relates to the duration of the gesture. With directing-to, the length of the gesture is usually quick and direct, or transitory (Clark, 2003, p. 262), possibly ending once the addressee has looked in the direction of the object or has confirmed that they see it. In contrast, placing-for deictic gestures tend to be longer, or continuing, because the object of reference is held in place at the indexing site.

When it comes to indicating multiple objects, directing-to is usually the more appropriate route, as pointing can easily direct the viewer's attention to many objects consecutively in a wide space. With placing-for, multiple objects over a wide space cannot be indicated because a placing-for requires the objects to be placeable at an indexing site.

The final difference between placing-for and directing-to is that placing-for is easier to retract than directing-to. Once the speaker is finished referring to the object, or if they want to change their reference, all they need to do is to remove the object from the field of view. With directing-to, however, once an object has been pointed to, the speaker has to verbalize or gesture in some way to let the interlocutor know that the indication is over. Table 2.1 provides a visual of some of the features of Clark’s

directing-to and placing-for categorizations. It will be important for the reader to keep these dimensions in mind for the remainder of the thesis as I will repeatedly make reference to them.

Placing-for	Directing-to
Joint accessibility of signal: “The object is accessible to everyone in a conversation for an extended period of time” (p. 262)	*Immovable objects:* “Objects that are difficult, impossible, or inappropriate to move or place” (p. 263)
Clarity of signal: “Continuing presence of an object makes it easy to resolve disputes about what is being indicated” (p. 262)	*Dispersed objects:* “Indicate objects one by one that are dispersed over a wide area” (p. 263)
Revocation of signal: If an object is placed for people to see and then removed, those viewing it can see that the object is no longer being referenced (p. 262)	*Directions:* “Indicate a direction - such as which way a car went” (p. 263)
Memory aid: The presence of the object is a constant reminder to the interlocutors on what it is that is being referenced and talked about (p. 262)	*Complex referents:* “I can point at a bottle of shampoo and, by saying ‘that company,’ refer to Procter & Gamble, the company that made it” (p. 263)
Preparation for next joint action: “Placement generally leaves the object in an optimal place for the next step in the joint activity” (p. 263)	*Precision timing*: “If I tell the baker, ‘I want that, that, and that, but not that,’ I must time my pointing to coincide with the right *that’*s” (p. 263)

Table 2.1: Dimensions along which placing-for and directing-to differ. (Clark, 2003, pp. 262-263).

Clark’s (2003) proposal that directing-to and placing-for generally differ in their use is empirically testable, still, his theories have never been tested. Therefore, one of the goals of my analysis will be to see to what extent this characterization of placing-for and directing-to carries over to a virtual environment.

Chapter 3: Methods

In this chapter, I focus on explaining the methodology that underpins this research. In 3.1 I discuss the selection criteria for the games and the tutorials and how it impacts the analytic procedures used. In 3.2 I discuss how qualitative, quantitative, and mixed methods are applied to the data. In 3.3 I examine the ethical decision making this type of research requires. Finally, in 3.4 I describe the system of coding that I designed for this research, and the issues I had to resolve while applying the coding to the videos.

3.1 Potential Games and Selection Criteria

A first step in the research design is to identify the sample of games to be studied. There are many genres of video games that I could choose from, the top ten of which are: action, sports, adventure, battle royale, role-playing (RPG), racing, fighting, real-time strategy, simulation and first-person shooter games (Straits Research, 2020). There are many tutorials for each of these genres. As the focus for this research is gestures, and micro-analytic in approach, sampling from all game tutorials is not possible. My goal instead was to gather a variety of gesture samples from completely different games in order to have a diverse sample that showcased the different methods content creators use to communicate in an online setting. This allowed me to investigate into how the affordances provided by the games or programs affected the deictic gestures the creators used. When I started the process, there was no guarantee that there would be any similarities or differences in gesture use when choosing different games, but from initial observation of the tutorials I realized there were deictic gestures that were common amongst many content creators. I also predicted that the different

types of games would provide different tools for the content creators to use, which, in theory meant that there would likely be a variety of gestures used.

The five games I initially selected represent a good microcosm of the gaming world because of the variety of genres they represent. *The Elder Scrolls Online* (*ESO*) and *Ark: Survival Evolved* are massively multiplayer online role-playing games (MMORPG or MMO for short), which is a sub-type of RPG. Although *ESO* and *Ark*: *Survival Evolved* are both MMO's, *ESO* requires the internet, and the game can include other players from all over the world. In contrast, *Ark: Survival Evolved* can be played offline, and allows for single player mode. On the other hand, *Saints Row IV* is an action-adventure game, *The Sims 4* is a simulation game, and *Starcraft II* is a real-time strategy game.

Since the games are from different genres, I predicted that the types of deictic gestures and communicative methods the content creators use for their respective YouTube tutorials would vary as well. For example, since *The Sims 4* is a simulation game that does not continue when the game is turned off, players can pause their game and return without missing any game events. This pause feature allows content creators to talk to their audience without having to worry about the game continuing without their supervision. For *ESO,* pausing the game is impossible since the game is an MMO that is continuously 'on'. If the player leaves, the game world continues, and the player will miss any events that occur within the game while they are away from it. Content creators who want to show certain aspects of *ESO* thus must find other methods by which to focus on certain aspects of their gameplay. They usually choose to record their

gameplay in some way and then add their explanation on top of the video through the editing capabilities of video editing software.

The selection of the tutorials was informed by my position as an experienced game-player while at the same time trying to balance several different factors to increase the range of games covered. In the first stage of data selection, there were a total of five games I considered for this research: *Ark: Survival Evolved, Saints Row IV, The Elder Scrolls Online (ESO), The Sims 4,* and *Starcraft II.* I chose these games because I was familiar with their mechanics and gameplay. Familiarity with the game is an important selection principle, as in-depth knowledge of the game is crucial for interpreting whether a cursor movement is communicative, as opposed to merely serving functions within the game world.

Having selected these five games, I began to watch a variety of tutorials about them. This viewing process allowed me to become familiar with the typical structure of a video tutorial and helped me to begin narrowing down the criteria that I wanted to focus on. Each game had many types of tutorials, by many different content creators. Possible tutorial types include, among others: playthroughs, reviews of the overall gaming experience, gameplay analysis, and game mechanics explanations. Not only can content creators choose from many different types of tutorials, but they can also choose from many different types of formats, some of the most extant being: using the mouse to reference content on the screen in conjunction with speech; having text, rather than speech, embedded on top of the video to indicate and communicate with the audience; and placing objects of reference in the audience's view in conjunction with speech or text. For some of the games (e.g., *Saints Row IV*) there were different types of

gameplay, one example being modes of gameplay such as levels of difficulty, e.g., beginner, intermediate and advanced, and the content creators would address these different modes by making multiple videos for the same game. The way players choose to play a particular game is nearly infinite and there are tutorials for many different occasions.

First and foremost, I decided that for a video to be considered for this research, it needed to contain spoken language and communicative mouse movements. Many content creators of game tutorials place a music track over their recorded video or only explain their actions using embedded text. In contrast, the videos I selected all had the content creator verbally explaining their actions in detail. Furthermore, I decided to focus on tutorial videos that explained to novice game players how to undertake particular actions in the game, such as how to pick up and place objects. I considered ‘novice players’ to be either those players who are new to the game or new to some component within the game. I expected that in tutorials for this intended audience, the content creator would use a greater number of deictic gestures, as the likely audience would not necessarily be familiar enough with game jargon to understand what the content creator was verbally referencing.

Tutorials were ruled out if the content creators did not play the game with the mouse as that would render it impossible to study the communicative functions of mouse cursor movements. This required that the tutorials I chose had the content creator playing the computer version of the games and not the console version (e.g., PlayStation, Xbox) that makes use of a game controller. Furthermore, I limited the tutorials to games that included the mouse cursor as an independent feature, i.e., the

actual mouse arrow cursor, visible on the screen, rather than include games, such as some first-person shooters, where there is no mouse cursor visible on the screen, and the use of the mouse directly manipulates integral parts of the game, such as the direction the character is facing.

Second, I decided to focus on the type of tutorials that taught about mechanics of the game while the content creators were playing it, and not the type of tutorial where the game settings or other material was presented. From my earlier observations of tutorial videos, I had noticed that these lesson segments had a more diverse sample of deictic gestures. It seemed that when the content creators are actually playing the game they relied more on movement and actions with the mouse to make themselves understood.

The final selection criterion for tutorials was the length of the video. Tutorial videos on YouTube vary in length from a couple of minutes to several hours, for example, gaming videos average around 24.7 minutes (Statista, 2021). A frame-by-frame analysis of gestures with the ELAN software is a time-consuming process, so I decided to limit the portion of the videos I would examine to those that met all of the prior criteria and that were between ten and thirty minutes in length as this was comparable in length to other papers on gesture research that analyzed video recordings. Such micro-analyses of gestures reported in the literature include truncated analyses, both in terms of time and in terms of the number of gestures. For instance, an entire dissertation revolving around only 1.2 seconds of gesturing in Hassemer's study (2009), to the analyses of gestures in Winter et al. (2013) which analyzed only 11 brief gestures in total, dealing with a cumulative 5.5 minutes of data. From this perspective,

the present analysis deals with substantially more data than is analyzed in other studies with a micro-analytic outlook on gesture. In dealing with 48 minutes of total video data, the current investigation is broader in scope than many other gesture analyses conducted in the field.

3.1.1 Game Selection

During the process of selecting tutorial videos, two of the five games I started out with were eliminated for a variety of different reasons. These games were: *Ark: Survival Evolved* and *Saints Row IV.* While in principle eligible given the stated criteria above in section 3.1, watching the tutorial videos revealed that the type of open world nature that is specific to *Ark: Survival Evolved* would make analysis more difficult. Typically, open world refers to a type of game where the players do not need to follow a linear, or pre-established, narrative set by the game developers, but rather wander the world of the game freely, completing pre-established ‘quests’ (an individual mission) and other tasks at their own leisure and in their own order. However, *Ark: Survival Evolved* does not have any set quests or any prescribed structure. Having no structure at all resulted in certain steps on how to play the game being skipped or omitted from the explanations of the tutorials, in contrast to content creators giving a full overview of a specific topic.

The second game that I considered but then eliminated was *Saints Row IV*, which is an open world action-adventure game. When searching for tutorials on the *Saints Row IV*, the videos were similar to play-throughs, which were about completing individual side quests or play-throughs of the main story. Play-throughs are videos where the audience simply watches the content creator play the game, rather than

receive instruction on *how* to play the game. As with *Ark: Survival Evolved*, there were not many mouse movements used because there was very little interaction with a user interface (UI) and movements of indication were generally not performed.

At the end of this selection process, I was left with three remaining games from the original five: *The Elder Scrolls Online, The Sims 4,* and *Starcraft II*. I was able to find tutorials for all three games that were for beginners and/or players new to the particular part of the game explored in the video. As I intended to focus on primarily the qualitative aspects of the data, I decided that a sample of such size would not be an issue because having “a single case (or event) or a small number of cases (or events)” (Gerring, 2017, p. 20) can be a benefit for qualitative studies. These three game videos were sufficient to provide a large pool of mouse cursor movements (see section 3.4) for in depth investigation while still limiting the study to manageable proportions.

Regarding the games themselves, there are very few similarities between the three games, other than that they can all be played with by a mouse. This is absolutely necessary as the goal of this research is to observe mouse cursor movements as a way to indicate. The other similarities found between the gaming tutorials, I chose, relate to the gestural data from the analysis discussed in the annotation section in 3.4 and in Chapter 5. In contrast, there are many differences between the three games. *ESO* is an online only game, whereas *Starcraft II* and *The Sims 4* do not require the internet to play. While *Starcraft II* can be played as a single player game, almost all gameplay is online in multiplayer mode against other people. For *The Sims 4*, going online merely provides access to content created by other players, such as pre-built houses or pre-made sims. *ESO* has both an open world and linear game option, however *The Sims 4*

and *Starcraft II* do not. *ESO* has many possible paths for a starting player to choose from when they enter the game. The role, the race, and the gear they choose all have an effect on the development of the gameplay. On the other hand, the starting path for the other two games is somewhat predetermined and is the same for all new players. In *The Sims 4,* a player will start with a family, either created or one that was saved in their files and will then place this family in an apartment or house in the sim world and start to play. In *Starcraft II,* the player will always start with a main base and then build up that base with other buildings and robots.

In summary, table (3.1) below provides an overview of the criteria used in selecting the games and tutorials.

The criteria used for choosing the games:

- Diversity of the games (based on genre and playstyle).
- The game represents one of the top 10 genres (no repeated genre).
- The game had to be played and primarily controlled via the mouse (e.g., no game controller or exclusive reliance on the keyboard).
- Experience with the game to facilitate interpretation of mouse cursor movements.
- Includes tutorials for beginners on YouTube.
- Games with a relatively structured goal-development pattern (e.g., type of pathway through games and therefore the kinds of instructions that might be given).

The criteria used for choosing the tutorials:

- Videos that feature spoken language (no text only or “Let’s Play” without commentary).
- Presence of gesture and speech.
- Tutorials that used a mouse to indicate and make references.
- Clarity of pronunciation.
- Focus on gameplay as opposed to other aspects of gaming. Preference for step-by-step instruction.
- A variety of observed gesturing behavior.

Table 3.1: A list of the criteria that was used in the game and tutorial decision making process.

3.1.2 Background of *ESO, The Sims 4,* and *Starcraft II*

There are an estimated 16.19 million players or subscribers worldwide for *The Elder Scrolls Online* and it was ranked 7th out of 130 MMOs (as of October 24, 2021) for player population (MMO population, n.d.). This game was developed by ZeniMax Online Studios and published by Bethesda Softworks in 2014. It is a fantasy game, where the setting is Tamriel, an open world continent where players can travel between the regions. Players can play solo or in groups ranging from two to twenty-four players in each group. There are a few types of content that players can choose from: quests, dungeons, and trials. Quests are goals a non-playable character (NPC) conveys to a player to complete. The quests are short stories that involve the NPCs in some way. Once the player completes the quest, they are usually given some type of reward. They do not always involve fighting, but often include finding an object or completing a task. Dungeons are small, contained areas or arenas that also feature a storyline for the players to complete. The main aspect is to fight the monsters contained within, which are often difficult and require multiple players to defeat. These dungeons are structured to allow up to four players to group up and participate together. Trials are essentially dungeons; but they are longer, more difficult, and can have up to twelve players in a group. The rewards from quests, dungeons, and trials are different from each other, thus compelling players to play all three.

The Sims 4 is a simulation game developed by Maxis and was published by Electronic Arts in 2014. *The Sims 4* is part of *The Sims* franchise which is one of the best-selling game franchise in history (Gee, 2007; Gee & Hayes, 2012). The fourth instalment is the most recent release from this franchise. Players control created

characters called ‘sims’ which are generally human characters. They can also create houses from an empty plot of land or buy pre-made houses with in-game money for the sim to live in. Players control their characters’ lives in all aspects, such as dating, careers, family, and satisfying their characters’ needs and desires.

Starcraft II was published by Blizzard Entertainment and came out in 2010, as the latest installment of the *Starcraft* franchise which first came out in 1998. The game can be played with other people or in single player mode. In this game, the players are military commanders of one of the three available races: protoss, zerg, and terrans. The goal is to defeat the enemy, usually by destroying their bases and buildings. Both games of the *Starcraft* franchise are also exercised as eSport (electronic sport), which describes games that are played competitively for prize money, with spectators viewing games of professional players live. *Starcraft II* was chosen because it is one of the most popular eSport real-time strategy games and is usually in the top twenty most viewed and played Esports. However, it is sometimes hard to pinpoint where it ranks at currently because the genre of Esports is in constant flux, and what might be the number one game this year may change by the next.

3.1.3 Tutorial Selection

The number of suitable tutorials I analyzed for each game varied depending on the results of the YouTube search. Something to keep in mind is that whenever a search is made on YouTube, the videos that are displayed will change and are different for different people, depending on their IP address, search history, etc. Searches on YouTube will differ even if the search text is verbatim from one day to the next on the

same computer that logged in from the same household. It is not possible to state how many tutorials exist for each game as videos are constantly being added or removed from YouTube. Furthermore, YouTube itself does not quantify the total number of videos it contains that matches search terms. The videos I chose were from those that were displayed on the day I originally conducted my search (August 19, 2018). If the same search terms are used in the future, it is possible that the videos I selected would still be found, as well as others that are more recent. To safeguard against link rot and ensure reproducibility of my research, the videos have been downloaded.

As I began the process of selecting YouTube tutorial videos for each of the three games, I started with *ESO*. I first searched on YouTube for tutorial videos on *ESO* and looked at some of the themes they shared relative to aspects of the game design. These included tutorials on the topics of: dungeons, trials, gold making, crafting armor or weapons, quests, and character builds. There were also videos made for beginners, but many of them did not fit the criteria set out above in section 3.1.1 (see table 3.1).

When selecting the topic to focus on from this game, I decided to remove gold making from the list because in the videos I observed there were not a lot of gestural indications and it did not really explain how to play the game. From the other five topics: dungeons, trials, crafting, quests and character builds, I gave each category a number and used a random number generator online to select one of them because I could not decide which topic to explore. The result of the number generation was the category of trials. However, the selection process was not yet finished. There were eight trials in the game at the start of this thesis that I could choose from. Although all of the trials had potential, I could not decide between them. Thus, in line with what others have done for

similar selection problems (e.g., Ferchaud, Grzeslo, Orme & LaGroue, 2018) I used a random number generator to make the decision for me. As a result, the trial called Cloudrest was chosen out of the eight and I began to search for a tutorial video specifically on that trial.

On YouTube, there were less than ten videos on Cloudrest. Many of these were play-throughs with music overlay and contained no instruction or explanation about the trial. At the time, T3hasiangod, who is from The United States, was the only content creator whose videos met all the criteria enumerated in section 3.1.1 (see table 3.1). He had created four guides on the four different boss fights found in this trial. The one I chose for this thesis was one I had already encountered in earlier stages of the content selection, and from those initial observations I knew it would provide me with rich data. T3hasiangod's video is 10:52 minutes long with 8,471 views (as of September 27, 2021).

For *The Sims 4*, only four of the first 40 videos that were displayed on YouTube satisfied the criteria laid out in table 3.1. Most of the videos that did not satisfy the criteria were based on how to construct buildings and houses in the game rather than how to play the game. From the remaining four videos three were eliminated for the following reasons: The first video was difficult to understand due to the speaker's pronunciation. The second video's content creator was difficult to follow, as they jumped around to different parts of the game rather than explaining the game step by step. Finally, the third video contained many parts, which would require a new player to watch more than one video to understand the gameplay. Deligracy, who is from Australia, was the only one whose video not only satisfied the criteria laid out in section 3.1.1 (see

table 3.1), but her video was succinct and clear in providing an explanation on how to play the game. Deligracy's tutorial was 24:20 minutes long with 1.8 million views (as of September 27, 2021).

Of the first 40 *Starcraft II* tutorials on YouTube there were nine tutorials that had potential to be chosen based on the selection criteria in section 3.1.1 (see table 3.1). I used a random number generator to select one tutorial from among those nine. I realized almost immediately that this was a mistake. Although all the videos satisfied the criteria laid out in section 3.1.1 (see table 3.1) from initial observations, there were some videos that were better than others at explaining how to play the game. I went through and watched the videos carefully. From the nine, I chose BeastyqtSC2, who is from Serbia, because his explanations were clear, organized, and easy to follow. BeastyqtSC2's video was 45:03 minutes long (with 160,063 views as of September 27, 2021). For this tutorial, I only transcribed and gathered data from the middle section of the video (total time: 13:35 minutes). This was done to keep it consistent with the other two tutorials and also because only this section contained actual gameplay, whereas the other sections were explanations of the game's settings.

3.1.4 Maximum Variation Sampling

Although the final selection criteria for each tutorial varied between the games, this is not a defect, but rather a necessity due to using different games, each of which have a different presence on YouTube. Since the games and types of tutorials are so different, this imbued the research with a wider pool of potentially different deictic gestures than might be visible if the games were similar in nature. Thus, the selection process is

broadly in line with maximum variation sampling, where having a small diverse sampling strengthens the research by ensuring that there is enough variation even in studies that use small sample sizes. As stated by Patton (2015):

> [W]hen selecting a small sample of great diversity, the data collection and analysis will yield two kinds of findings: (1) high-quality, detailed descriptions of each case, which are useful for documenting uniquenesses and diversity, and (2) important shared patterns that cut across cases and derive their significance from having emerged out of heterogeneity (p. 283).

A benefit of maximum variation sampling is that a researcher can gain insights into a specific topic without necessarily having a large sample size (e.g., Johannes, Vuorre & Przybylski, 2021). When sampling from a large spectrum of data, one of the goals is “to be as representative as possible” (Glen, 2016, p. 2). However, I chose to use a micro-analytic approach, focusing on the detailed frame-by-frame analysis of specific gestures in context (Carter, Gibbs & Arnold 2012; Winter, Perlman & Matlock, 2013). This meant that it would be difficult to examine a large body of data (as micro-analysis of gesture data is inherently time-consuming). Therefore, a researcher can use maximum variation sampling to achieve diversity while maintaining a workable sample size (Hull, Brunelle, Prescott & Sargent, 2014; Malliet, 2007). For example, Čulig, Katavić, Kuček and Matković (2014) used this method “to include as wide as possible range of gamers, mutually differing in genre preferences and gaming intensities, in order to obtain diverse gaming experiences” (p. 67). By having a diverse sample, they

were able to obtain different perspectives of what gameplay meant to the players and their motivation for playing the games they played.

3.2 Mixed Methods

The decision on whether to use the qualitative method, quantitative method, or a mixed method which combines both, depends on the type of research, the needs and intent of the researcher, and the research inquiries. Qualitative methods are often used to provide depth of understanding, underlying meaning, and interpretation of patterns and data (Angouri, 2018; McKay, 2009; Palinkas et al., 2015). Qualitative research establishes or explores theories behind why things happen and deals with detail with a concern for context. Qualitative analysis allows the researcher to gain information from interviewing or questioning individuals to gain first-hand access to that person(s) way of thinking or feeling (Edley & Litosseliti, 2018). In contrast, quantitative methods involve numerical data that is generally collected for large sample sizes, with researchers hoping to ascertain results that are as unbiased as possible, and that can generalize to a larger population (Ivankova & Creswell, 2009). It can be used to explore trends (Baker, 2018), and it focuses on "*how much* or *how many* there is/are of a particular characteristic or item" (Rasinger, 2018, p. 119).

One of the limitations of quantitative analysis is that the "data does not provide evidence for *why* populations think, feel, or act in certain ways" (Goertzen, 2017, p. 13). However, combining quantitative methods with qualitative ones helps alleviate this limitation to some extent, as the *why* can often be informed from a detailed analyses of contexts. On the other hand, a limitation of purely qualitative research is that there is no

formal demonstration that any obtained results are generalizable to a larger population. Though both qualitative and quantitative analysis have their own limitations, the advantages of each complement the other as “both...provide a depth and breadth that a single approach may lack by itself” (Ivankova & Creswell, 2009, p. 136).

This view informed my decision to use a mixed method approach which can be broadly defined as “research in which the investigator collects and analyzes data, integrates the findings, and draws inferences using both qualitative and quantitative approaches in a single study or a program of inquiry. A key concept in this definition is integration” (Tashakkori & Creswell, 2007, p. 4). Although the qualitative and quantitative methods “are usually oriented toward somewhat different goals...they nonetheless inform each other in a useful fashion” (Gerring, 2017, p. 28). The two concepts are not fused together to make one inseparable piece, but rather each method can be considered a separate piece that works independently, with its own purpose, while simultaneously being able to work together to form a more complete whole. As stated by Angouri (2018), “mixed methods designs can contribute to a better understanding of the various phenomena under investigation compared to their exclusively QUAL/QUAN counterparts; while quantitative research is useful towards generalizing research findings...qualitative approaches are particularly valuable in providing in-depth, rich data” (p. 41). All of these quotes speak to the benefits of using a mixed methods approach, which is why I decided to use a mixed methods approach to take advantage of the benefits each of the two methods provides.

Though this thesis incorporates a mixed methods approach, both quantitative and qualitative methods are not used equally, rather there is an emphasis on a

qualitative approach, because it has been suggested to be “particularly appropriate when studying a topic which has been little explored…” (Ivankova & Creswell, 2009, p. 140). Since up to this point, there has been no extensive analysis of mouse cursor movements in video games from the perspective of gesture research, my thesis deals with a relatively under-scrutinized subject that requires a detailed exploratory look at novel data. Qualitative analysis thus provides the best way to describe what is happening and “a rigorous and efficient way to answer…‘how’ and ‘why’ questions” (Hamilton & Finley, 2020, p. 2).

In this study, I wanted to understand ‘how’ content creators communicate with their audience in the video game tutorial context, especially when traditional methods of indication, such as pointing with an actual finger in the physical world, are not, and cannot, be used. In order to know ‘how’ the content creators communicated with their audience, I first needed to know ‘what’ they used to communicate, and this resulted in a research methodology that developed in different stages. The first stage was qualitative and exploratory, looking at every single mouse cursor movement that occurred in the selected videos and deciding which ones of these movements acted as communicative devices, e.g., deictic gestures, and whether they interacted with deictic expressions. After I had determined ‘what’ to study, I used quantitative methods in the second stage to count how many deictic gestures were used by each content creator. I then split those deictic gestures and expressions into their different types and used those numbers to extrapolate broad patterns that held across the content creators’ work. This gave me a focused direction in ‘what’ to look at so that I could examine each type deeply in a qualitative fashion and discover ‘how’ it was being used, such as whether or

not a link existed between the gestures found in my data and with existing frameworks of gesture analysis. In other words, quantitative methods “determine the prevalence of themes” (Patton, 2015, p. 657) and “uncover behaviors and trends” (Goertzen, 2017, p. 12), while qualitative methods are used to explain the behaviors and themes.

3.3 Ethics

Collecting data on user-generated content sites like YouTube, raises important ethical concerns. As with all applied linguistic research, a central consideration to the collection of the data is to “avoid harm to yourself and to others” (Page, Barton, Unger & Zappavigna, 2014, p. 58) and to make sure of “the safety of all involved in the research” (University of Birmingham (UoB) Code of Practice, 2021, p. 8). With the development and expansion of virtual environments, e.g., social media, public forums, and gaming communities, there have been debates on the ethical practice’s researchers must consider and incorporate in their research. Some examples of ethical concerns that need to be considered are the risk factors to both the researcher and the subjects, protecting confidentiality of participants, ascertaining what is considered public or private, and deciding in what situations and to what extent participant consent is necessary. These matters have been debated widely in the applied linguistic research which takes new media data as its focus (Facca, Smith, Shelley, Lizotte & Donelle, 2020; Gelinas, Pierce, Winkler, Cohen, Lynch & Bierer, 2017; Gupta, 2017; Hokke, Hackworth, Quin, Bennetts, Win, Nicholson, Zion, Lucke, Keyzer & Crawford, 2018; Kitchin, 2003; Legewie & Nassauer, 2018; Sugiura, Wiles & Pope, 2017). I will discuss how they apply to my research in the following sections.

3.3.1 Confidentiality and Anonymization

Confidentiality and anonymization are two of the most fundamental elements that every researcher implements when working with human subjects, as it is our job as researchers to “maintain...confidentiality in order to achieve protection of Intellectual Property Rights, commercially sensitive data, personal data and other confidential information” (University of Birmingham Code of Practice, 2020, p. 8). Confidentiality is the act of keeping something (e.g., data) secret, secure and undisclosed. It can be difficult to determine to what extent confidentiality is needed for YouTube videos that have been made for the public. The users themselves in a way have removed confidentiality from their own work by making it accessible to anyone with an internet connection, and they include information they are comfortable sharing, since they can always edit their videos before uploading them. If a content creator does not want their video to be seen, they can make it private. However, research that focuses on YouTube videos, may very well require the researcher to download the videos. For example, transcription programs, such as ELAN (discussed in section 3.4), requires the video file to be accessible from the computer.

One caveat to consider is that although the content creators may have agreed to share their content with the public, they may not have consented to it being used for research purposes. Therefore, even if the YouTube video is made publicly, the downloaded version should be kept confidential, as something that was perceived private at one point can become public later, or vice versa (British Psychological Society (BPS), 2021; Whiteman, 2012). For this thesis, I provided confidentiality for the downloaded videos by excluding the downloaded versions in the publication of this

work. Instead, I provide the link to each of the videos in the references. This is so that if the authors do decide to remove their work or make it private, they are given complete control of accessibility. The downloaded copies of the tutorials are kept on password protected hard drives, stored in a locked drawer. In special cases where I must share the videos, e.g., to my PhD supervisors, they can access the videos only after being provided a private link from google drive. No one without the link has access to the videos or research content. After the completion of this thesis, all data and videos will be removed from google drive.

Anonymizing personal information ties in with the overarching goal of ensuring the safety of participants, as it is one way to prevent participants from being exposed and potentially harmed. Anonymization means that even in the event confidentiality is lost, information is still protected because it is not easily linked to real identities. One of the benefits of the online setting, is that there is self-anonymization, in which individuals anonymize themselves by using a username or internet handle. One can argue that this self-anonymization is enough to protect the identities of individuals since there may be no way to tie their internet handles to their legal identity. In addition, the name, photo and any information they may put in an online profile might not be accurate or true, thus contributing to the difficulties of finding out who they are. One conundrum researchers may face is whether or not they should still anonymize the pseudonyms, usernames, or self-anonymized names of the individuals (Buchanan & Ess, 2008). Even though the username the participant has chosen may not be their legal name, and there may not be a way to identify them, it is still possible to find that person's posts, account, or profile online through search engines. For example, searching for a YouTube content creator

by their username in YouTube's search engine will populate the search results with that creator's videos.

The factors mentioned above are part of the reasons that the degree of anonymization of online materials varies on a case-by-case basis. There are cases of online research where anonymization may not be necessary. YouTube content creators are generally seen as an example of this, especially those whose accounts are public and aimed towards a viewing audience. Some YouTube content creators who wish to distance their legal identity from their content, choose to use a pseudonym. There are also some content creators who use their real names and include themselves talking or acting in their videos (e.g., Mark Rober). These content creators may not care if people know who they are, or it may be a way for them to gain recognition for their work in a way that is directly tied to their physical world identity. In fact, anonymization by a researcher in these instances may be counter to the wishes of the content creator. By its very nature the process of anonymization removes credit from a creator. One of the primary goals for content creators on YouTube is to increase their viewership. The more views they receive, the more chances they have of monetizing their videos, such as through paid memberships by their viewers, or by having the audience watch advertisements before viewing the video content. By anonymizing the content creator a researcher potentially not only deprives them of the credit that they deserve for their work but they also decrease the views they might have garnered thanks to broader exposure from the publication of a scholarly article.

Because of the reasons stated in the previous paragraph, for my research I decided not to re-anonymize the names of the content creators so that they can receive

recognition for their work. In any case, the research project was low risk to the content creators. There was no information about the content creators or any identifiable data, aside from what is in their YouTube video, which is game content, and their internet handles, which are public. These internet handles, which are a form of self-anonymization, are what I use to reference them. I do not know their real names or any information about them, and for the purpose of this research, it is neither necessary nor helpful for me to know this since I am only observing and analyzing how they communicate with their virtual audience through their YouTube tutorials. I conducted no interviews and did not try to gather their perspectives or their opinions.

The next criterion I considered when deciding whether I could use the tutorials I chose was whether they were considered public or private videos, and under what conditions decided this.

3.3.2 Public versus Private

An ethical challenge for those interested in researching information found online is teasing apart the puzzle of what is considered public and what is considered private (BAAL, 2021). Much of the internet is public, a collection of data that can be accessed by anyone with an internet connection (Walther, 2002) and does not require membership or registration of any type to view the contents. On the contrary, a private environment is not available to the public, and has “access…restricted to the creator of the content and his or her invited guests” (Sveningsson Elm, 2008, p. 75). In private contexts, a website usually requires individuals to apply for membership in order to see the content. The differentiation between public and private is not always a clear

dichotomy thus it is important to keep in mind that the concept of public and private “are constructed rather than innate” where the “meaning and significance of these terms need to be worked through in reference to specific locations/contexts rather than simply applied as if they were self-evident or natural” (Whiteman, 2012, p. 51).

YouTube is an example of both a public and private platform. YouTube provides their users with a private option, where the content creator can upload content privately for their own use or for the use of those they choose to share it with. Of course, YouTube provides many videos that have been made publicly for anyone to watch as there is a “lack of restricted entrance” (Rosenberg, 2010, p. 24) e.g., no need for membership. During the decision-making process of the videos, I used for this thesis, there were a few ethical factors that were necessary for me to consider.

The right to access the tutorials I used was to some extent determined by whether or not they were considered to be publicly available. First and foremost, they were all uploaded onto a public forum that is accessible to anyone with the internet and YouTube. The videos were searchable through YouTube’s search engine, and they were found within those results. Videos that are private do not appear. Second, in the case of accessing the tutorials for this thesis, there were no gatekeepers, which allow or deny individuals access to certain content or locations (Whiteman, 2012). YouTube does not require an individual to sign up for an account to access the public videos, and there are no limitations placed by YouTube that disallows individuals to have access. The only possible gatekeeper would be the privatization of the videos by content creators, not the company of YouTube. In addition, the videos I selected could be determined as public because under YouTube terms of service, content creators:

> grant each other user of the Service a worldwide, non-exclusive, royalty-free license to access your Content through the Service, and to use that Content, including to reproduce, distribute, prepare derivative works, display, and perform it, only as enabled by a feature of the Service (such as video playback or embeds). For clarity, this license does not grant any rights or permissions for a user to make use of your Content independent of the Service. (YouTube Terms of Service, Effective as of March 17, 2021)

The content that uploaders post on YouTube are for public use of other users, as YouTube is a "forum for people to connect, inform, and inspire others across the globe, and acts as a distribution platform for original content creators…" (YouTube Terms of Service). However, there are certain legal conditions such as copyright that need to be followed. This can be seen with the caveat "[f]or clarity, this license does not grant any rights or permissions for a user to make use of your Content independent of the Service" at the end of the block quote above. The way the material is presented in this thesis falls under the purview of the service agreement and YouTube's purpose. I give credit to the content creators of the videos I am using, and this research project in no way reproduces or redistributes the content, neither does it claim the videos as mine. I only analyze the publicly available materials provided by YouTube.

The last condition that influenced my decision that the tutorials I used were for public consumption was when the content creators directly addressed an open-ended and non-specific audience, using words such as 'you'. Although there were indicators that showed that the videos I chose were made for public consumption, thus making it

likely that I do not need to receive consent (Sixsmith & Murray, 2001; Walther, 2002; see also Tri-Council Policy Statement (TCPS), 2018 and BPS, 2021), I nonetheless took the additional step of securing informed consent.

3.3.3 Consent

Informed consent is “the process by which researchers can allow participants to negotiate, document and agree [to] their contribution to a research project” (Page et al., 2014, p. 71). When it comes to online research, there are varying beliefs amongst researchers as to when it is required for them to seek consent (Whiteman, 2012). One belief that most ethical bodies of research seem to agree on is that consent and ethical regulations should be considered on a case-by-case basis as not all research has the same risks and ethical conditions (The British Association for Applied Linguistics (BAAL), 2021; Markham, 2012; Page et al., 2014; Whiteman, 2012).

The ideas discussed above in section 3.3.2 become particularly pertinent, in that the interplay between what is regarded as private and public plays a part in determining the necessity of receiving consent. Many researchers believe that data “that is drawn from publicly available materials...naturalistic observation” (Kitchin, 2003, p. 414) or “behavior performed within the public domain[,] may be observed and researched without consent” (Sixsmith & Murray, 2001, p. 425). If a researcher is analyzing a publicly published video as just an observer of what is happening in the video, and the results and discussion have no involvement or ties related to the person making it, e.g., no personal data or confidential information, and the research poses little to no risk, then consent or an ethics review may not be necessary (UoB Ethics of Practice, 2021),

so long as the researcher can justify their choices. On the contrary, if a researcher interviews a content creator of a YouTube video and examines the comments made by the viewers, such as in Potts (2015), then an ethics review, consent and/or anonymization may be necessary. In the case of the research presented here, I chose to ascertain informed consent, if only out of respect for the content creators, and to make them aware that the content they chose to make publicly available was being studied for research purposes. Moreover, ascertaining consent of the content creators removes any doubt about the permissibility of using the content, thereby satisfying researchers with different opinions about whether consent should be sought in a situation such as that in the following paragraph (see also Whiteman, 2012).

I tried to contact each content creator of the tutorials I collected data from through various means because I did not know which contact method would reach them or if any would. For example, for the content creator T3hasiangod, I posted a comment under his tutorial on YouTube, and I messaged him on Discord, a program that allows people to talk or text in real-time online, since he had posted about a particular group he was a member of that I was also familiar with. For BeastyqtSC2, the content creator of the tutorial on *Starcraft II*, I emailed him and sent him a private message through Twitter. For Deligracy, the content creator of *The Sims 4* tutorial, I first visited her Discord channel that is open to the public, and I privately messaged a moderator who provided me with an email to contact Deligracy. Thanks to these attempts I was able to reach each of the content creators and was able to secure their informed consent to use their tutorials in my thesis (see Appendix A-C).

In summation, the ethical principles that govern research are highly dependent on the content of that research, there is no concrete rule. Instead researchers need to adapt their ethical principles to focus on “contextually situated ethical practice where the actions of the researcher are informed by the specific nature of their research and research settings (rather than involving the application of a general principle such as ‘informed consent should always be obtained’, ‘participants should always be anonymised’, etc.)” (Whiteman, 2012, p. 9). Not only should that be kept in mind but also;

> “there is no reason to place greater burdens on researchers using the Internet unless some aspect of their study – other than that it happens to be using the Internet – calls for it…to require all Internet researchers either to acquire signed informed consent, or to be asked to demonstrate the real ages and competence levels of subjects, especially in a minimal to-zero harm project, in ways that are not required of alternative methodologies, seems unreasonable.” (Walther, 2002, p. 213).

3.4 Annotations and ELAN

3.4.1 Software

The program used for transcribing and organizing the data was ELAN. ELAN is “versatile enough to accommodate different fields of research. Whereas some tools are specialized…for field linguistics or audio-only speech analysis…ELAN has always been a multipurpose, multimodal annotation tool” (Lausberg & Sloetjes, 2009, p. 847). I chose to transcribe using ELAN primarily because it is a free, open source, publicly accessible

program that anyone can use, and this makes it possible for others to replicate my process of annotation without having to obtain possibly expensive proprietary licenses.

Another advantage of ELAN is that it permits multiple layers of transcribing, allowing for more than just one type of annotation to be applied to a certain part of the video (see Figure 3.1). ELAN provides users multiple options when working with the recordings, such as zooming and screenshotting. ELAN is also a widely used program in the gesture research community (e.g., Cooperrider, Fenlon, Keane, Brentari & Goldin-Meadow, 2021). There are other transcribing programs such as MAXQDA, NVIVO (see Cserzo, 2019), and ATLAS.ti, all of which are used for video, audio and document coding and analysis. However, these are all programs that need to be paid for, and using paid software is generally seen as being against core open science and reproducible research principles (Winter, 2019), restricting access to annotations via expensive licenses and making it impossible for everybody to reproduce analyses without paying for software. As recommended by Winter (2019) in the context of linguistics, “whenever possible, use software that is accessible to everybody” (p. 48).

In addition to ELAN, there are at least two other transcription programs that are free: EXMARaLDA and TranscriberAG. EXMARaLDA and TranscriberAG are focused on the transcription and annotation of audio and video files, with EXMARaLDA also having a tool for managing corpora and analysis. Although these two programs are also free and open-source, ELAN was also chosen for two other reasons. ELAN is the de facto standard for gesture analysis in the gesture research community (Colletta, Guidetti, Capirci, Cristilli, Demir, Kunene-Nicolas & Levine, 2015; Cooperrider, 2014; Holler, Kendrick & Levinson, 2017; Rosborough, 2016). The second reason I chose

ELAN for module three was based on familiarity. I had used ELAN in module two, therefore, for this thesis I decided to continue using this program.

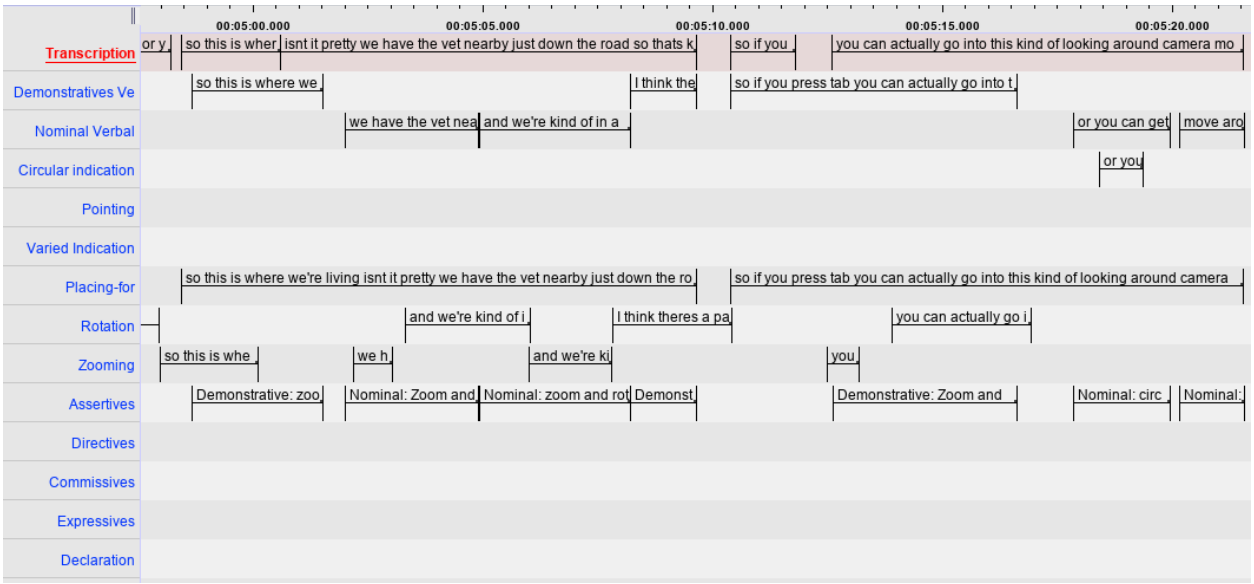

Figure 3.1: Sample tier taken from Deligracy's ELAN file. The category titles are on the left side. The time stamp can be seen on the top ruler, and the annotations I made are in the center.

3.4.2 Annotation

Before I started to code the speech and the movements that were used in the videos, I first had to transcribe all of them into separate tiers. Having different tiers was a way for me to tease out all the components of the tutorials that were deictic gestures or expressions. My focus in this study is on how deictic gestures and expressions are used together, and by annotating the deictic gestures and expressions on separate tiers, I could look at a bigger picture and see where they coincided with each other.

First, I created a transcription tier which included all the speech in the video. From there, I examined which verbal utterances were used to make a reference and coded them as such. I found that the verbal utterances that were coded as a reference tended to occur as two types: the use of demonstratives or nominals (see Figure 3.1).

How I determined whether a word was used as a deictic utterance was based on the situation and context they were used in. The ones that were deictic expressions always referenced something. For example, in Figure 3.2, at the beginning of the video when Deligracy is creating her sim she says, “I don’t really like any of these poses” (Deligracy, 1:14). Using the context, the ‘these’ references the different poses of her sim because she is flipping through different images of her sim standing, and her actions with the deictic gesture lines up with the deictic expression.

Figure 3.2 Deligracy: Deligracy is going through different poses of her sim (top figures). The poses are displayed when players choose the world they want to live in (bottom figure). 1:06.

I then created two tiers: one to contain the demonstratives, the other to contain the nominals. The demonstratives tier included only referential proximals and distals such as “here”, “this”, and “that”, whereas the nominals tier contained the nouns, or the name of the object of reference. The final piece of linguistic analysis I did was to add five tiers for the five different types of speech acts: assertives, directives, commissives, expressives, and declaratives. Although there are many speech acts contained within the game tutorials, my interest is in how content creators use deictic behaviors to communicate with their audience. Because of this I only analyzed the speech acts that contained deictic expressions and gestures. As one speech act can contain multiple deictic expressions and gestures, including demonstratives, nominals, and gestures, for each speech act I labeled how many of each were present.

The next aspect of the transcription process was to create a tier that contained all of the deictic gestures. To determine what to put in this tier, I watched the videos and wrote down each type of deictic gesture I saw and how each content creator used mouse movements to indicate. I determined which types of mouse movements were produced with communicative intentions based on the context, i.e., not with the primary function of propelling the game world further (*goal-oriented movement* - Novack et al., 2016) but with the primary function of establishing deixis. For example, I knew a deictic gesture was used when the content creator took their mouse and pointed at something on the screen and then proceeded to discuss it, and these indications were often used with either a demonstrative or nominal. From my observations and notes, I realized that there were some mouse movements that all three creators used: pointing, circular pointing, placing-for, and varied indications, a catch-all category for indications that

were neither a circle nor a point; this last one will be explained further below. My creation of categories was very straightforward and was based on the structure of the mouse movement. If a mouse movement was a direct point with the mouse, it would go under pointing, and if it was circular in structure it would go under circular pointing. This also applied to the mouse movements that were specific to each content creator, as I will discuss later, that were presumably driven by the affordances of each game's user interface, such as: 'rotation' and 'zooming' for Deligracy, 'boxing' for BeastyqtSC2 and 'stopping-for' for T3hasiangod.

The gesture annotation system was partially informed based on already pre-established terminology, such as 'pointing', a word used across much literature on gesture, and placing-for (Clark, 2003). For example, a direct point with the mouse in the virtual environment was similar to a direct point with the finger in the physical world. In line with Clark's analysis as points being actions that direct attention, such a mouse cursor movement clearly drew attention to the object it was indicating. The other annotations were informed by the data presented in this thesis. Since I found some types of deictic gestures that did not have physical world counterparts already established in previous research, a personalized coding system was created, which involved the creation of new labels for particular types of gestures. The labeling for deictic gestures was based on what they looked like when they were being used, e.g., circular indication.

Some instances of the coding were ambiguous, or uncertain, and required further categorization. For example, some mouse movements had no consistent pattern or were not used consistently throughout the tutorials. Therefore, I decided to create a

category that was titled ‘varied indication’, and this was for any mouse movement that was used to indicate that did not take the form of a direct mouse point, a circular mouse movement, or occur more than once with the same form. This general category was created because from initial observations some mouse movement did not fall into any ascertainable category that is currently present in the literature and there was often no common pattern for some of these indicating mouse movements. In addition, there were not that many incidences, so I did not create individual categories for each of the forms, but rather one category to encompass all anomalies.

The coding ‘rotation’ and ‘zooming’ were labeled as such for Deligracy because that is what they represented. Although rotation and zooming are features seen in the physical world, such as a video camera rotating and zooming, it is not considered a gesture of indication. The movement and actions of rotation and zooming in *The Sims 4* is similar to the way rotation and zooming are used in the physical world in form, however, they are used as a way to indicate in the game and tutorial. Boxing was coined as such because when BeastyqtSC2 used this mouse movement, it took the form of a green box shape. Stopping-for was a hybrid of the word ‘stop’ and Clark’s (2003) terminology ‘placing-for’. In the case of T3hasiangod, the ‘stopping-for’ use had the same function as a placing-for, but instead of the video being played continuously, the content creator stopped the video to indicate the content that was stopped on. Although further detailed discussion of the deictic gestures studied in this paper will be given in Chapter 4, it is first imperative to show visual examples of these deictic gestures. I have split them into two sets of visuals: first are the deictic gestures that are

shared by all the examined content creators, second are those that are unique to particular content creators.

<u>Shared Deictic Gestures:</u>

Circular Indication:

Figure 3.3 Deligracy: Deligracy is making a reference to the Sim. 1:01.

Pointing:

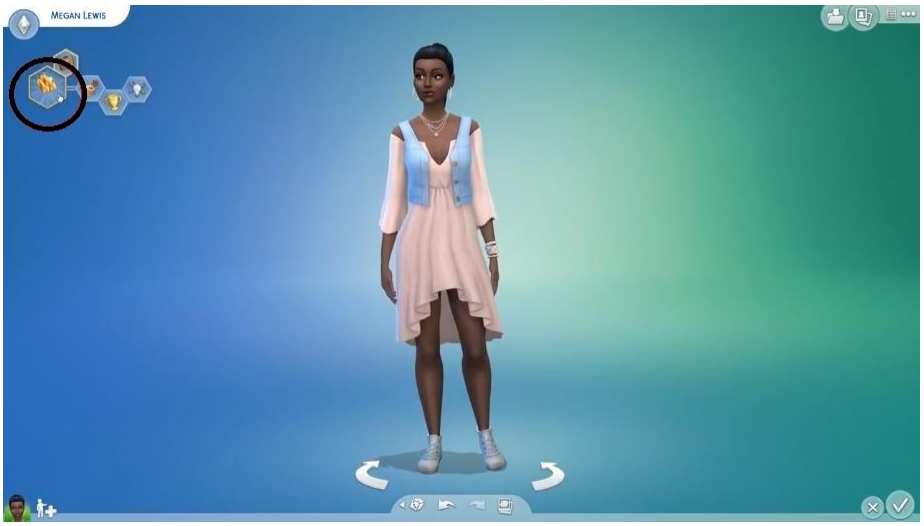

Figure 3.4 Deligracy: Deligracy uses the mouse to directly point to the object in the upper left corner. 0:45.

Varied Indication:

Figure 3.5 BeastyqtSC2: BeastyqtSC2 is indicating an area he is referencing. The form of the mouse movement does not fall into any specific category. 18:27.

Placing-for:

Figure 3.6 T3hasiangod: T3hasiangod is referring to the large tentacle in the image, however he does not stop the video when he is talking about it. 1:24.

<u>Unique to Content Creator Deictic Gestures:</u>

Boxing:

Figure 3.7 BeastyqtSC2: BeastyqtSC2 uses boxing to reference a location on the screen. The large green box is the outline of the box he made in the game, and the small yellow box is the mouse pointer. In this case he started the click and drag at the upper left corner of the green box then dragged the mouse diagonally down to where it is currently situated in the image. 19:33.

Rotation:

Figure 3.8 Deligracy: Deligracy moves her mouse from right to left to rotate her character. 0:40.

Zooming:

Figure 3.9 Deligracy: Deligracy zooms in to her character. It's hard to see but by looking at her feet placement with the arrows, it can be seen in the second image her feet are below the screen because of the zooming in. 0:18.

Stopping-for:

Figure 3.10 Deligracy: T3hasiangod stops his video and references the small yellow circle on the bottom of the screen, where the players are standing (I drew the black arrow to indicate to the small clear yellow circle he is referencing). He pauses the video at 3:58 and starts the explanation. The image above was taken at 4:35.

After annotating each type of deictic gesture and expression, I categorized each type as a directing-to or placing-for, based on Clarks (2003) existing research on the distinction. Directing-to deictic gestures were any that directed a listener's attention to a reference, whereas in placing-for situations, an object of reference is placed in front of the listener. Pointing, circular pointing, varied indication, and boxing, were categorized as directing-to, whereas those labeled as placing-for: stopping-for and zooming and rotation were categorized under placing-for. Chapter 4 on typology discusses these distinctions in more detail.

Chapter 4: Typology

The typology presented here is one that I have developed to contribute to the growing research on gesture. Although the categories contained herein are a result of my own analysis, they were influenced by the work that others have done on gestures in the physical world, such as Clark's (2003) dichotomy of directing-to and placing-for. This section presents some of the results of my thesis, namely, a qualitative discussion of the different types of gestures that occur in the YouTube gaming tutorials examined. The emphasis here is on looking at the different categories that I developed for my coding scheme. Section 4.1 will introduce my typology for directing-to and section 4.2 will discuss the different types of placing-for found in YouTube gaming tutorials. Chapter 5 will build on what is discussed here.

4.1 Directing-to

In the virtual environment, directing-to can take a number of forms of which the following are some examples:

1. Circular indication (Figure 4.1)
2. Pointing with the mouse (Figure 4.2)
3. Using an in-game tool to indicate, e.g., fireball (Figure 4.4)
4. Using the mouse to indicate by boxing an area of the screen (Figure 4.5 & 4.6)
5. Varied indication: using the mouse to indicate by creating shapes other than boxing and circular indication (Figure 4.7)

In the context of gaming tutorials, speakers use the mouse in multiple ways to communicate with the audience by directing their attention to the object of reference. Evidence that mouse movements communicate in this way have been suggested by previous studies. Chen, Anderson, and Sohn (2001) investigated the correlation

between eye movements in relation to mouse movements on web browsing. They found that of the regions the mouse cursor visited (or moved to) “84% of them were also visited by an eye gaze...among the regions that the eye gaze didn’t visit, 88% of them were not visited by the mouse cursor” (p. 282). If the eyes follow the mouse movement to the new location, one could say the viewer’s focus is directed from where it was previously to whatever is being indicated. If that is the case, it can be argued that when a content creator moves the mouse to something on the screen, it is highly likely that those watching will shift their gaze to the location or object that the mouse cursor is referencing.

The first type of directing-to that will be discussed is circular indication. Circular indication is a type of gesture categorized as a directing-to because it draws the viewer’s attention to an immovable object. It also provides clarity of signal (Clark, 2003) as to which object is being referenced. There is no real physical world equivalent to circular pointing because typically a simple point can be used to indicate an object. In the virtual environment, even though circular gesturing is not precise in that it is not a direct point, it can provide more salience than a pointing gesture in some contexts. It can be used as a way to reduce ambiguity by grabbing and directing the viewer's attention to something that is often difficult to pick out. Since it is a constantly moving gesture, it is highly visible and that makes it easier to see the reference. This type of gesture was used by content creators to reference when they needed to pick out one object from among many, or when one object takes up a great deal of space on the screen.

These aspects of circular indication are best illustrated by Figure 4.1 where

BeastyqtSC2 is explaining building placements and how the player can have the different main structures share another small structure (labeled ‘Add-on’ in the figure). The two main structures are called barracks. Currently BeastyqtSC2 has two barracks placed, however he wants to explain that if one was a factory and the other were a barracks, their location on the screen can be changed accordingly in order for both to have access to the add-on in the figure. Since he only has two barracks and no factory, he tells the viewers to pretend that one of the barracks is a factory. By circling it a couple of times, not only is the creator directing attention to the barracks, but he is removing ambiguity as to which of the structures he is referencing. He is emphasizing which of the two structures, that look the same, is the ‘pretend’ factory.

This example is also a type of conceptual blending, which “refers to a set of cognitive operations for combining (or blending) words, images, and ideas in a network of ‘mental spaces’ to create meaning” (Nordquist, 2019 [Fauconnier & Turner, 2003]). He asks the audience to pretend as if the barracks is a factory, and by doing this he is blending the first mental space, which is what is seen on the screen (the two barracks) with the second hypothetical space (the factory). By doing this, he prompts the audience to create a new mental image, which is for one of the barracks to be the factory.

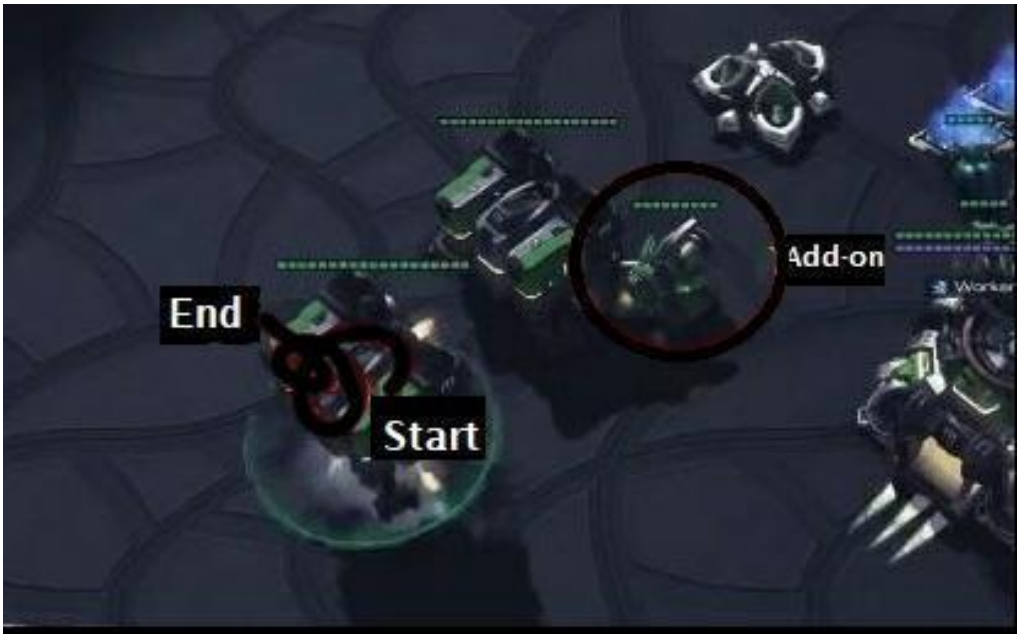

Figure 4.1 BeastyqtSC2: "...you can swap your add-ons. So lets say this barracks is a factory…" 27:01. The circular gesturing with the mouse indicates the pseudo factory. The circle highlights where the add-on is. The circle and lines were drawn by me.

The next type of directing-to that will be discussed is pointing. I categorize this deictic behavior as directing-to because it fulfills the feature of directing the viewer's attention to an immovable object as established by Clark's system found in table 2.1 in Chapter 2. It also illustrates the feature of clarity of signal because it picks out which object is being referenced clearly, differentiating it from other objects in the area. Finally, in the virtual environment, it serves as a reminder to the audience of what they should be focusing on (Clark's dimension: Memory aid). This dimension according to Clark is a feature of placing-for in the physical world. However, in the virtual environment it can be both. In the physical world the manual point is pointed in the direction of the object, and rarely put on the object itself. In the virtual environment, the mouse pointer is always placed right on top of the object of reference, thus making it more salient. Figure 4.2 is an example showing that a directing-to gesture can have features of a placing-for.

Figure 4.2 is an illustration of how pointing is used in YouTube gaming tutorials. At the beginning of the video, Deligracy explains the various icons on the screen while gesturing to each of them. On the bottom of the screen there is a bar. When a player hovers on a certain location on that bar, a pop up displaying more details appears. This specific pop up shows the different types of seasons and which one the player is currently in. The pointing gesture the content creator uses indicates the location of her verbal utterance, “the time, the season, the weather.” The verbal cues signal to the viewer that the content creator is going to point to something, and the gesture directly points out what the viewer should look at.

Figure 4.2 Deligracy: “...we’ve got the time, the season, the weather…” 5:54. The box highlights the pop up that shows up when the mouse hovers over the long icon on the bottom of the screen. The circle highlights where the mouse currently is pointing to on the screen. The box and circle highlighting the objects were drawn by me.

There are some differences between pointing in the physical world and in the virtual environment. In the physical world, if a reference is made to an object in a

cluttered environment and the interlocutor does not understand what is being referenced, the speaker may have the option of going to and touching the object. However, in the virtual environment, a direct point is already placed on top of the object of reference. The speaker does not have the option of getting closer to the object in order to provide clarity. In cluttered environments, such as in Figure 4.3 below, salience is lost because finding the mouse among all the visual clutter can be quite difficult. In these types of situations pointing is not salient, but rather ambiguous. To counteract the ambiguity, circular gesturing is sometimes used to provide further clarity as was hinted at above.

Figure 4.3 BeastyqtSC2: Starcraft II: Ambiguity - Where is the mouse? 21:11.

The next type of directing-to that will be discussed is that of using an in-game tool to indicate an object. This one differs from the other types as it depends heavily on the affordances provided within the game. The content creators will make use of these

affordances to communicate with their audiences but there is no real consistency between them as each program and game contains different in-game tools. For example, in Figure 4.4, the game *The Elder Scrolls Online* allows players to be mages, or magic users. Magic users can use a staff (the item the character is holding in Figure 4.4) that shoots fire. In this game, one way to indicate is to use the staff to shoot a fireball at an object. In Chapter 2 I briefly mention tool-assisted pointing, such as using a pen to indicate. In a sense the staff in this game is an extension of the avatar. Since the avatar does not have the ability to point with their hand, as it is not programmed into the game, the game player must use another method to indicate. In this case the staff and fireball are an extension used to indicate. The fireball makes use of two of Clark's features: clarity of signal and direction. Clarity of signal occurs from the fireball being bright and visible. This grabs the audience's attention to what is happening. When the fireball is sent out of the staff, it is a feature of Clark's direction because the content creator shoots the fireball in a specific direction, which directs the viewer's attention in the same direction towards the referenced object.

Figure 4.4 The Elder Scrolls Online: An avatar using a fire staff to indicate a lamp post. This was a screenshot taken from my game.

The next directing-to gesture that will be discussed was distinct to a single YouTube gaming tutorial. BeastyqtSC2 used a gesture that was found neither in T3hasiangod or Deligracy's videos, nor in the physical world. This gesture often took on the form of what I call 'boxing'. In this category, the user can click anywhere on the screen and drag the mouse, creating a box shape over any object or the background. I classified it as a directing-to gesture because it was used to indicate immovable or stationary objects, such as the background of the game. Boxing was also considered directing-to because it fulfilled the feature of indicating many objects or locations that were dispersed across a wide area. Most interestingly, it illustrates how clarity of signal in the virtual environment is a feature of directing-to as well as placing-for. This is seen in Figure 4.5 (the first boxing picture) when BeastyqtSC2 indicates locations that are not

only far from each other but also immovable. He first marks one area with a boxing gesture (left picture) then marks another area (right picture). The boxing makes it clear to the audience what he is referencing, and the boxing coinciding with the “here” makes his indications precise.

Figure 4.5 BeastyqtSC2: “...put a building like here, or here…” 29:58. The box around the boxing gesture was drawn by me. It outlines the green box that is in the game and created by the mouse pointer. The first “here” corresponds to the first image, and the second “here” corresponds to the second image.

Boxing is a particularly useful gesture because it decreases ambiguity in cluttered backgrounds, much like circular pointing does. In BeastyqtSC2’s video, there are often many things on his screen and his pointing gestures tend to disappear into the background. Much like in the discussion of circular pointing above, BeastyqtSC2 has many immovable and dispersed objects across his screen, and he cannot simply approach the referenced object and touch it, nor can he pick it up and bring it into the view of the audience. BeastyqtSC2 has to somehow direct the attention of the viewer to the single referent among many that he is discussing. Boxing is the method by which he accomplishes this task. This gesture is used to reduce ambiguity and make the object of

reference visible. The form of the boxing gesture is a bright green color that can be any size, which can attract the attention of his audience.

Figure 4.6 most clearly illustrates how boxing can be used to decrease ambiguity. In this example BeastyqtSC2 is telling the audience that as a Protoss player, they need to take three probes and place them on the gas structure. Boxing here is used in two different ways, the first being a tool used to highlight the probes that he wants to draw attention to, and the second as a way to select the probes. He drags his cursor across the probes, thus creating a highlighted box that selects them. In this example, the content creator is combining the two ways boxing is used, by simultaneously selecting and highlighting the probe. The deictic gesture of highlighting accompanies the speech and indicates to the viewer what a probe is by calling attention to it.

Figure 4.6 BeastyqtSC2: “So first things first is, if you’re a Protoss player, you need three probes on the gas.” 17:00. The boxing in the figure coincides with the words “you need”. The box I drew highlights where on the screen the boxing is located, and the shape is a square to represent how the boxing shape looks.

Boxing in the game is similar to the click and drag on the computer, which allows the user to select multiple objects at one time by selecting everything within that box. In cases such as Figure 4.6, boxing is both a functional and communicative device. In this figure, BeastyqtSC2 is talking about probes, or the little robots, in general. There was no need to specifically point out which probe should go on the gas, just any three probes. In this situation, he used a larger gesture of boxing to select however many can fit into the box just to indicate to the audience what probes are. In this way it also seems to decrease the amount of effort needed to indicate because finding a specific probe out of dozens of similar ones and then clicking on it may require more mental energy and time. The content creator, by using boxing, can simplify or decrease the amount of “internal computation” needed to reference (Kirsh 1995, p. 31).

The final type of directing-to that will be explored is categorized under varied indications, which consists of gestures that can take on almost any shape other than a circular indication or a boxing. All of the gestures in this category took on one or more of the features of directing-to found in Clark’s (2003) categorization system. An example of a varied indication is Figure 4.7, which illustrates how they work in the context of the game *Starcraft II*. In this game, robots (SCV) bring resources they have gathered from a source location to the main base. In this example there are SCVs on both the mineral patch and the gas line. Since BeastyqtSC2 requires SCVs to build new structures, he has to take them from one of these resources and give them the new task of building a structure. He is explaining which SCVs to take from which resource hub, in this case, the ones from the mineral patch, not the gas line. He uses a varied indication gesture, outlined in Figure 4.7, to provide a clear signal and indication towards the location of the

entire immovable mineral patch area.

One interesting facet in Figure 4.7 is that when the content creator says “mineral” there is a long pause, and an “um” before he says the word “patch”. The pause in speech and gesturing seems to happen because he is trying to recall the word “patch”. Previous research (Ahlsen, 2015; Kistner, Dipper & Marshall, 2019; Kita, 2000; Krauss, et al., 2000) has found that spontaneous speech gestures can be used “as part of the retrieval process” (Krauss, 1998, p. 55) and that seems to be the case here. The gesturing of the mouse during the long pause after the word “mineral” is fulfilling two purposes: indicating the mineral patch and participating in the word retrieval process.

Figure 4.7 BeastyqtSC2: “So whenever you’re building buildings with Terran, make sure to pull the one or two, however many SCV’s, from the mineral patch and not the gas...” 25:55. Lines were drawn by me.

Gestures that had iconic features were also placed into the varied indications category. Iconic gestures are so called because they either imitate the way an object

looks, or they are shaped by what the speaker thinks the object looks like. For example, a speaker could say “go upstairs” and perform a circular rotating gesture with their index finger going upwards. In their mind, the idea of a staircase is that of a spiral staircase. In contrast, another person could use the index and middle finger to represent a person walking up stairs that are not circular. While the virtual environment does not provide the three-dimensional space needed to do such manual gestures as seen in these examples, it does offer one clear advantage for iconic gestures: the objects they are referencing are right there on the screen, therefore they can just outline the way the object looks, they do not need to project an image from their mind and use gestures to represent the image.

Iconic gestures in the videos were usually categorized as varied indication and placed under directing-to because they often performed some of the functions of directing-to such as providing clarity of signal or directing attention to immovable objects. However, it was sometimes unclear if a gesture should be placed under varied indications or another category because the content creator’s intent could not be determined. A clear example of this uncertainty is Figure 4.8. Here T3hasiangod references a circular object by making a circular gesture around it. Is the circular motion used because the content creator is following the circular form of the object while indicating, in which case it should be classified as an iconic gesture? Or is he using a circular motion to indicate as he does in other instances where the object of indication does not happen to be a circle? Such instances seem to combine iconicity with deixis, but it is difficult to determine the intention of the content creator (see Kendon, 2004, p. 202 for a similar example in gesture that combines iconicity with deixis).

Figure 4.8 T3hasiangod: "...you can see there is an AoE that grows underneath whoever is targeted by this..." 4:01. The lines were drawn by me.

Figure 4.9 is another instance that shows how iconicity and deixis can be blended. In this example, Deligracy is explaining the needs panel of *The Sims 4* and how the player must pay attention to this panel: if a particular need goes into the red, negative consequences will happen. When she is explaining the consequences, she moves her cursor back and forth on the energy bar, twice. The first time coincides with "...if their energy goes to red..." and the second coincides with "...and collapse…" Starting from the ZZ circular icon under the energy panel, she moves the cursor right then back left, and holds the cursor point on the ZZ icon representing the lowest point on the bar.

This gesture could be considered iconic because she is gesturing horizontally along the bar, mimicking the way it is shaped (horizontal) and the way it moves when

the need decreases (movement toward the left) or increases (movement toward the right). Even though the viewer does not see red, when she holds the cursor on the ZZ, the viewers understand that if the green bar goes down to the ZZ (the left side of the bar), the bar will become red because of her iconic movement in conjunction with her speech. The second time she moves the cursor back and forth, she is not only reiterating the dangers of the need going down, but she is also using this gesture to emphasize the dire consequences if it is ignored by elucidating exactly what will happen to the sim.

Figure 4.9 Deligracy: "...if their energy goes to red they will pass out and collapse" 8:03. The line, drawn by me, represents the movement of the mouse.

4.2 Placing-for

In the virtual environment placing-for can take on a few forms.

1. Generic placing-for (Figure 4.10 & 4.11)
2. Zooming and rotation (Figure 4.12 & 4.13)
3. Using a single object to indicate a site (Figure 4.14)
4. Embedding - images or text put on a pre-recorded video (Figure 4.15)
5. Stopping-for (Figure 4.16)

One way in which virtual environment placing-for differs from physical world placing-for is that YouTube gaming tutorials themselves can be argued to be a type of placing-for, which I am calling generic placing-for. The content creators make a video and then place it onto a video hosting site that allows anyone to click and watch it. The videos themselves are for the viewers' attention, and the viewers follow along with what the content creator says. The video is providing a constant source of joint accessibility so that the creator and the viewer are both focused on the same thing, albeit asynchronously. The video can be played, paused, and rewatched many times, therefore, it provides a constant source of clarity as to what is being indicated and it serves as a memory aid to the viewer. It even fulfills the feature of revocation for placing-for since at any time the content creator could choose to remove it from the video hosting site.

Since everything the user does occurs on the screen, one could contend that instead of directing-to and placing-for being clearly separate entities in the virtual environment, directing-to serves as a component within the placing-for structure. The video itself is a placing-for, constantly there for the audience to see. Anything that shows up on the screen can be indicated using other directing-to gestures such as pointing, varied indication, circular indication, and boxing. The viewer sees the content creator placing, moving, or deleting objects in general but there were also times when the content creator directed the viewers' attention to something specific they were doing on the screen simply by speaking about it. The viewer's attention is already focused on what is happening in the video, in a general sense of placing-for, but when the content creator specifically highlights something they are doing with a directing-to, the viewer is

expected to shift their attention to that specific reference within the video context. An example of this is in Figure 4.10. When Deligracy talks about the needs of the sim, she draws the viewers' attention to the needs panel by first zooming into that section of the video (placing-for). She then uses an iconic indication (directing-to) with the mouse, to point to the needs bar on top of the placing-for. In this segment she is explaining to viewers who may not have played the game before, or who know very little, how the needs bars work. She tells the viewers in the video that the needs panel is something important for them to pay attention to, and her gestures coincide with her speech.

Figure 4.10 Deligracy: "...if their energy goes to red they will pass out and collapse" 8:03. The line, drawn by me, represents the movement of the mouse.

Another generic placing-for that includes an example of directing-to is in BeastyqtSC2's *Starcraft II* video (see Figure 4.11). In this video there are many things happening, and the viewers see everything that is happening on the screen while he is explaining how to play the game. When he wants to talk about something specific or wants to direct the viewers' attention to something he is indicating, he may use the boxing tool as a directing-to, within the placing-for of the screen. Without indications, explaining gaming content would be extremely difficult because so much is happening

on the screen. Not only would it be difficult for the viewer to discern the specific object being referenced, but if they were to misunderstand and take a phrase out of context, it could have multiple meanings and result in a gameplay failure. Kendon (2000) argued that “speakers often employ gesture in such a way as to make something that is being said more precise or complete” (p. 51) and Figures 4.10 and 4.11 illustrate how the content creators adapted to the virtual environment to create virtual gestures that allowed them to communicate their meaning more effectively.

Figure 4.11 BeastyqtSC2: “So first things first is, if you’re a Protoss player, you need three probes on the gas.” 17:00. The box I drew highlights where on the screen the boxing is located and the shape is a square to represent how the boxing shape looks.

The next type of placing-for to be discussed is one that occurs within the generic placing-for context of the video: a placing-for within a placing-for. This is accomplished by manipulating the game itself in order to bring objects of reference that would otherwise not be visible into the view of the audience through such tools as rotation and zooming. These types of gesture fall under placing-for as they contain Clark’s (2003)

feature of joint accessibility of signal and their constant placement serves as a memory aid.

These features of zooming and rotation are best seen in Figure 4.12. Deligracy is explaining how to use some of the special mouse functions and is illustrating the rotating and zooming features by manipulating the screen. Although these actions perform a function in the game, they can also be considered deictic force because they change what is on the screen and this forces the viewer to look at what she wants them to. Rather than using the mouse to indicate, Deligracy occasionally uses these features to place an object of reference in the view of the audience or take the audience's view to the object of reference.

Figure 4.12 Deligracy: Zooming: "...scroll the wheel in and out to move in and out…" 4:37.

Rotation: "... press the wheel in and shuffle the mouse left and right to move it like this to pan…" 4:41.

Aspects of zooming and rotation are also clearly illustrated in Figure 4.13 where the content creator is showing the audience the neighborhood her sim will be living in. She references the area surrounding her sims's property, such as the nearby park and the vet clinic. When she is telling the audience about the neighborhood, she has to zoom or rotate her screen to show the audience what she is talking about, as these objects cannot be picked up and brought into the audience's view. In the physical world, this is accomplished by pointing to large objects such as buildings and cars. However, in the virtual environment, the affordances a game or program provides can do something beyond what is available in the physical world and bring these kinds of large objects into the view of the audience as a placing-for.

Figure 4.13 Deligracy: "We have the vet nearby…" 5:02. The circles around the building were drawn by me to indicate the building and are not part of the game.

The next type of placing-for that will be explored is that of using a placed object to indicate a site or space. I label these types of examples as a placing-for because the content creator is using the object as a guide or placeholder to indicate the environment around it, i.e., the indexing site, not the object itself. The indexing site has always been there because it is part of the game, the creator is simply “place[ing] an object or action for their addressees at a specialized site as a directive to interpret the object or action appropriate to that site” (Clark, 2005, p. 510).

The object that is being used to indicate provides a clarity of signal (Clark, 2003) to the viewers by delineating the size and extent of the referenced site. The object being used is not the end goal, but merely the means by which to direct the viewers’ attention. Once content creators are finished using the object as a way to indicate, they can remove it from the screen which signals to the audience that they are done.

In *Starcraft II* there is an affordance that allows an object that must be placed somewhere to be selected and then manipulated on the screen before being placed in its final location. When this occurs the mouse cursor changes and a silhouette of the object that was selected takes its place. This gives the player an idea of the size and look of the object before placing it down. In Figure 4.14, BeastyqtSC2 is giving an explanation and demonstration on how the base should be built and the things to consider when building it. Here he is telling the viewers not to block the back of the base because if any enemies come to the back of the base, it will be hard for his military units to move around.

BeastyqtSC2 uses a building silhouette that has not yet been built as a visual stand-in for hypothetical buildings that could be built. The building silhouette he uses in

the game emphasizes his verbal utterance of, “...try not to block the back of your base [with structures]...” He then takes the structures silhouette that has replaced his mouse cursor and makes a large circle around the current main base and resource nodes. This gives the viewers a visual of the area he is warning them against filling. He is using deictic force to get the audience to pay attention to an extensive area of the screen, the area that he has demarcated using a placed silhouette.

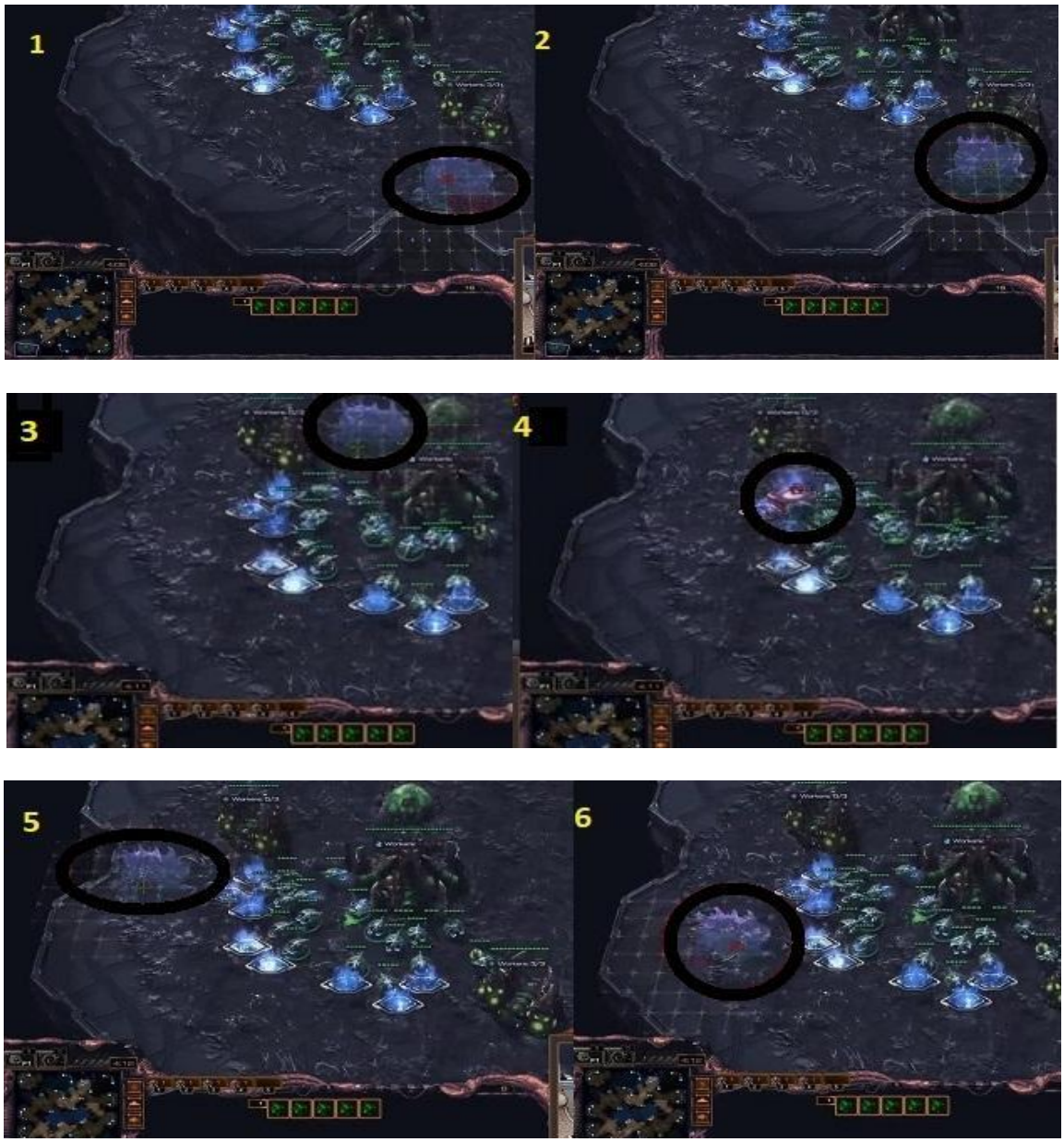

Figure 4.14 BeastyqtSC2: "...try not to block the back of your base too much. On some of the maps you can kinda almost fully wall-off your backside with your buildings…" 29:31. The circles in the images show the location of the mouse movements. The last image shows the path the mouse moved. The circle and lines were drawn by me.

The next two types of placing-for to be discussed both occur when the content creator manipulates the video recording, rather than the game. There are times when playing a game where the player cannot stop or pause what they are doing. In cases such as these, a content creator can pre-record their gameplay, then edit and narrate over it. There are at least two ways content creators make references to objects in such a pre-recorded video. The first way is through embedding text or images onto the pre-recorded video through an editing program (see Choi & Behm-Morawitz, 2017) in order to indicate or explain something. Figure 4.15 gives an example of placing-for where

embedded text is placed on top of the video footage of the gaming tutorials. The text in white “More of these orb dots with each higher difficulty” was written on top of the video by JerBearthe3rd, who is a YouTube content creator for *The Elder Scrolls Online*. In this example, his text is about a mechanic of the fight against an enemy. With the text, he uses an arrow to indicate the purple orb on the screen that he is referring to. In this figure, the embedded text is placed on top of the video, where it is accessible for everyone to see.

Figure 4.15 JerBearthe3rd: “It’s generally good practice to have just one DPS focus them on hardmode” 2:54. The content creator places embedded text to indicate an object on the screen.

A second method content creators use when referencing an object in a pre-recorded video is what I call a stopping-for. In a stopping-for, the content creator pauses the pre-recorded video through video programs such as Windows Media Player or QuickTime, and then proceeds to talk about what is on the screen in detail. Like other deictic gestures, stopping-for could potentially be analyzed as either directing-to or placing-for. However, it more clearly belongs to the latter because of the following four

dimensions from Clark's system: joint accessibility of signal, clarity of signal, memory aid, and revocation of signal. The pause on the screen makes the object being referred to visually accessible to the audience as the object is held still for the period of time that the screen is paused. This makes it clear to the audience what is being referenced, while also providing a constant memory aid. When the speaker is finished with their discussion, they can revoke the object of reference, in this case the stopped footage, by continuing the video.

The aspects of stopping-for are most clearly illustrated in Figure 4.16 from T3hasiangod. In this case, after the camera is turned to where the gryphon landed, the content creator pauses the video and begins to explain what is happening. T3hasiangod chooses to stop the video to explain the fight rather than speaking over the duration of the battle while it is playing. This seems to be because this section of the fight with the gryphon was too short for him to convey all of the information that he felt was important. The fight with the gryphon is 5 seconds long (6:43-6:48 minutes:seconds) if he played the video without any pauses. However, he pauses the video at 6:04 and talks for 39 seconds (until 6:43). By pausing the video, not only can the viewer focus on the image of the gryphon while it is being referenced by the content creator, but the content creator creates enough time to explain the referenced creature's features to his satisfaction.

Figure 4.16 T3hasiangod: "...so the gryphon comes back down...you attack the gryphon while it's on the ground then while it's in the air you attack the rider." 6:04. Video is paused.

This chapter provided a taxonomy of the various ways content creators use deictic gestures to indicate referents on their screen in the virtual environment of YouTube gaming tutorials and illustrated how the features of Clarks' (2003) dichotomy developed for the physical world are present in the virtual environment, albeit slightly differently. In his dichotomy, the features are separated into either directing-to or placing-for, however, in the virtual environment a feature, such as joint accessibility of signal, can apply to both a directing-to and placing-for gestures. Not only do some of the features apply to both types of gestures, but there are some deictic gestures that can only occur in the virtual environment, such as boxing. The content creators find creative ways of using what is available to them, such as programs or in-game tools to communicate with their audience. The following chapter will elaborate on how three content creators of different YouTube gaming tutorials use the affordances of their mediated context to indicate and reference objects in their tutorials.

Chapter 5: Results and Discussion

This chapter presents the results of the data analysis on the three YouTube gaming tutorials selected for this thesis. Section 5.1 provides quantitative data regarding which deictic gestures were used to establish deixis, and how often they were used in each of the YouTube gaming tutorials. Section 5.2 discusses the types of speech acts that co-occurred with gestures. Section 5.3 examines examples from the data that illustrate the different types of deictic gestures. Section 5.4 introduces Clark's temporal dimensions of directing-to and placing-for gestures to show how they apply to the deictic gestures in the YouTube gaming tutorials. Section 5.5 introduces proximal and distal demonstrative use with deictic gestures and their frequency. Further elaborating section 5.5, section 5.6 discusses how different factors, such as ambiguity, possibly influenced the choice in certain proximal and distal usage. Lastly, section 5.7 discusses the strategies that the content creators used to signify that mouse cursor movements in the virtual environment have communicative functions, i.e., what factors make a mouse cursor movement a communicative gesture.

5.1 Comparison of Deictic Gestures Across All Three Content Creators

The graphs and tables in this section report the number of times each gesture type occurred in each of the games as well as their percentages. First, I will present general information about which gesture types appear across all three games, followed by a more detailed investigation and explanation about the deictic gestures of each individual content creator, starting with *T3hasiangod*, then *Deligracy,* and finally *BeastyqtSC2*.

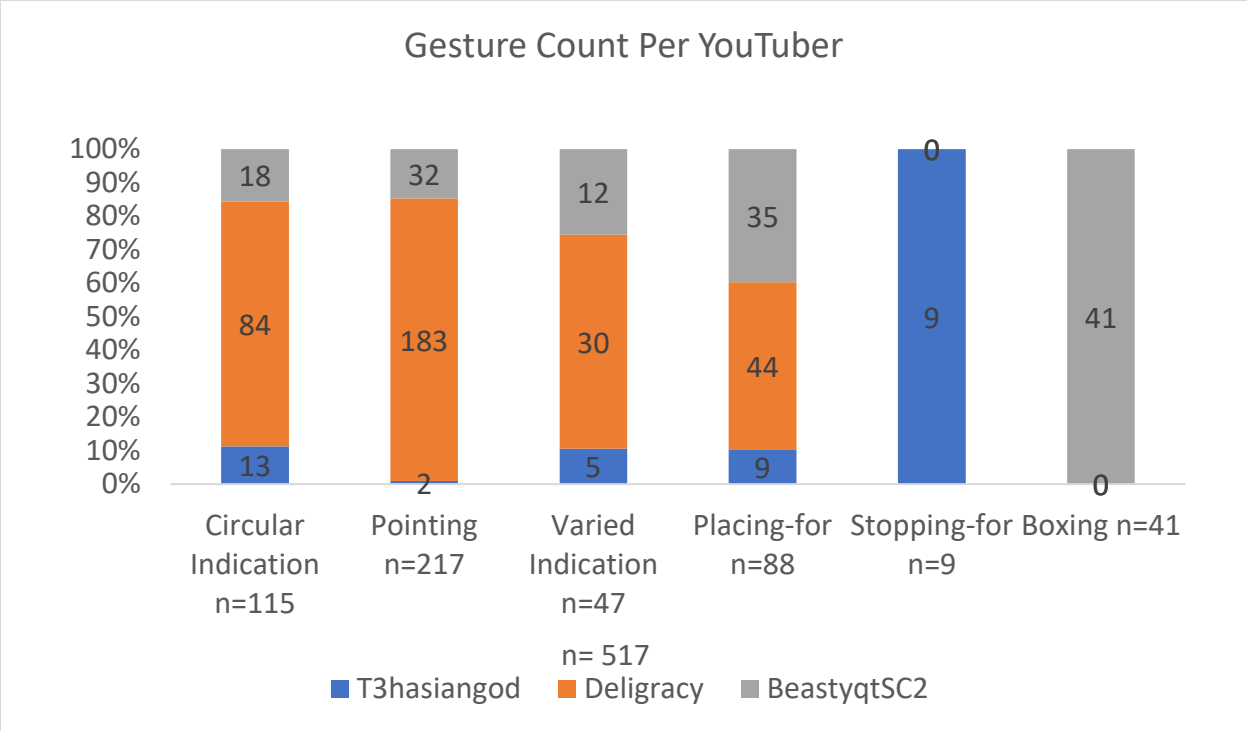

Figure 5.1: A comparison of the total count and percentage of each deictic gesture for all three content creators. This chart illustrates all gestures, regardless of linguistic context, used with both verbal demonstrative and nominal deictic expressions.

	Circular indication	**Pointing**	**Varied indication**	**Placing-for**	**Stopping-for**	**Boxing**
T3hasiangod	13 (34%)	2 (5%)	5 (13%)	9 (24%)	9 (24%)	0 (0%)
Deligracy	84 (25%)	183 (54%)	30 (9%)	44 (13%)	0 (0%)	0 (0%)
BeastyqtSC2	18 (13%)	32 (23%)	12 (9%)	35 (25%)	0 (0%)	41 (30%)
	115 (22%)	**217 (42%)**	**47 (9%)**	**88 (17%)**	**9 (2%)**	**41 (8%)**
517 (100%)						

Table 5.1: Counts and percentages of the deictic gestures for all three content creators. Corresponds to Figure 5.1.[1] Percentages are based on column-wise totals.

[1] There is a discrepancy in the number of deictic gestures and deictic expressions. Sometimes there was more than one deictic expression that occurred with one deictic gesture. There were also instances where more than one deictic gesture occurred with only one deictic expression.

The following gestures were used in the YouTube gaming tutorials to establish deixis: Circular indication, pointing, varied indication, placing-for, stopping-for and boxing. Table 5.1 shows the total number of occurrences for each gesture type per content creator. There were a total of 517 deictic gestures across all three videos. The most frequent deictic gestures were pointing, occurring a total of 217 times, which constituted 42% of all gestures. This was followed by circular indication, of which there were 115 occurrences in total (22% of all gestures). Placing-for had 88 tokens (17%), varied indication had 47 occurrences (9%), boxing had 41 tokens (8%), and stopping-for occurred a total of nine times (1.7%).

Instances of varied indication occur across all three videos but were rarely used in each of them. As for boxing, it was found only in BeastyqtSC2's video. Stopping-for had the least number of occurrences, which seems to be accounted for solely by the difference in how a content creator made their video: whether they were playing the game as they discussed it or if they were discussing a videotape of a prior gameplay session. Stopping-for was found only in T3hasiangod's tutorial, in which the content creator was discussing a video of a previous gameplay.

A comparison across content creators suggests there were similarities and differences between their use of deictic gestures. Overall, pointing was the most frequent. However, the overarching picture is one of diversity, with each content creator having a distinct gesture type that they used the most frequently. This attests to how rich the semiotic toolkit for establishing deixis is in a virtual environment: there are many different types of gestures, and the choice seems to be driven by individual differences in the content creators and the affordances of the game. In T3hasiangod's video,

circular indication occurred the most. Pointing was the most used in Deligracy's tutorial, and boxing occurred the most in BeastyqtSC2's tutorial.

In Deligracy and BeastyqtSC2's YouTube gaming tutorial, their origo is situated in the same temporal location and time as the game footage they are discussing because they are currently playing the game. However, for T3hasiangod, the verbal track is provided at a different temporal point, which is after the game footage had already been recorded. He took the footage of his gameplay and then talked over it explaining what he had done when he was playing the game in the past. This created a distance between him and the game as it unfolded and as he was interacting with a recording, and not the game itself, it was impossible for him to directly manipulate the gameplay while he was speaking. Other than this major factor between the content creators YouTube gaming tutorials, there is no obvious reason to explain why the different content creators had different patterns of gesture behavior in their videos. In order to understand why these patterns occur, we need to explore how the differently mediated contexts of each content creator affected how gestures were used.

5.2 Speech Acts that Co-occurred with Gestures

The tables in this section report the number of times each gesture type co-occurred with each type of speech act. First, I will present general data about which gesture co-occurred with each speech act across all three games. This is followed by a more detailed exploration of the speech acts that co-occurred with deictic gestures for each individual content creator, starting with, *T3hasiangod* then *Deligracy,* and finally *BeastyqtSC2*.

Assertive Speech Acts							
	Circular indication	**Pointing**	**Varied indication**	**Placing-for**	**Stopping-for**	**Boxing**	**TOTAL**
T3hasiangod	10 (33%)	1 (3%)	1 (3%)	8 (27%)	10 (33%)	0 (0%)	**30**
Deligracy	25 (22%)	54 (48%)	9 (8%)	24 (21%)	0 (0%)	0 (0%)	**112**
BeastyqtSC2	3 (9%)	9 (27%)	1 (3%)	3 (9%)	0 (0%)	17 (52%)	**33**
Directive Speech Acts							
	Circular indication	**Pointing**	**Varied indication**	**Placing-for**	**Stopping-for**	**Boxing**	**TOTAL**
T3hasiangod	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	**0**
Deligracy	1 (33%)	0 (0%)	2 (67%)	0 (0%)	0 (0%)	0 (0%)	**3**
BeastyqtSC2	3 (17%)	4 (22%)	2 (11%)	4 (22%)	0 (0%)	5 (28%)	**18**
Expressive Speech Acts							
	Circular indication	**Pointing**	**Varied indication**	**Placing-for**	**Stopping-for**	**Boxing**	**TOTAL**
T3hasiangod	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	**0**
Deligracy	1 (5%)	8 (36%)	1 (5%)	12 (55%)	0 (0%)	0 (0%)	**22**
BeastyqtSC2	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	**0**

Table 5.2: This table compares deictic gestures that co-occur with speech acts across the content creators. It is separated into each type of speech act that occurred; assertive, directive and expressive. The raw numbers indicate how many of each type of gesture co-occurred with each type of speech act. It is further divided by each content creator to show how many of the deictic gesture and speech act combination occurred in each tutorial. Percentages are based on row-wise totals.

Although there are five types of speech acts, only three were present in the data. There were a total of 218 deictic gestures that co-occurred with these speech acts. Assertive speech acts occurred the most across all three content creators as well as in each individual tutorial. T3hasiangod had a total of 30 deictic gestures that co-occurred with assertives. Deligracy had a total of 112 and BeastyqtSC2 had 33. Directives that

co-occurred with deictic gestures were only present in Deligracy and BeastyqtSC2's tutorials. Deligracy had three counts and BeastqtSC2 had 18 counts. Expressives that co-occurred with gestures were only present in Deligracy's video, and there was a total of 22.

Given that the YouTube gaming tutorials are designed to help the viewer achieve particular goals, the co-occurrence of gestures with speech acts is particularly important. As introduced by McNeill's (1992) seminal work, gesture and speech are an integrated unified system where both are "aspects of a single underlying process" (p. 23). This concept also applies to the utterances of speech acts and how they are used with deictic gestures. In the following section, the tables present the total number of deictic gestures that co-occurred with the speech acts used by each of the content creators. The speech acts that incorporated demonstratives in the utterance will be given special focus as the demonstratives seem to function as the link between the speech acts and their context. The tables below present the total number of deictic gestures that occurred in concurrence with the speech acts used by each of the content creators.

5.2.1 T3hasiangod's Speech Acts

	Assertives	Expressives	Directives
Circular indication	10 (33%)	0 (0%)	0 (0%)
Pointing	1 (3%)	0 (0%)	0 (0%)
Varied indication	1 (3%)	0 (0%)	0 (0%)
Placing-for	8 (27%)	0 (0%)	0 (0%)
Stopping-for	10 (33%)	0 (0%)	0 (0%)
TOTAL	**30**	**0**	**0**

Table 5.3 T3hasiangod: Counts for the speech acts that co-occurred with deictic gestures. Percentages are based on column-wise totals.

In T3hasiangod's gaming tutorial, only assertive speech acts were present. There was a total of 30 assertive speech acts that occurred. The two most frequently used gestures that co-occurred with this speech act were circular gestures and stopping-for, both with ten counts and contributing 33% each to the total amount of deictic gestures that occurred with assertive speech acts. Assertive speech acts that occurred with placing-for had a total of eight counts, contributing 27% of the total. This was followed by pointing and varied indication both with one count, each making up 3% of the total.

Stopping-for was one of the two most used deictic gestures that occurred with assertive speech acts. One possible explanation is that they took place in the part of the tutorial where T3hasiangod's focus was on explaining the mechanics of a boss fight. During this section, he would stop the video and state facts about what the boss would do while providing visuals of the mechanic he is discussing (Figure 5.2). It seems that he would do this because it was easier for the audience to understand his explanation

of the content when the normally fast paced fight was stopped.

Figure 5.2 contains an assertive that co-occurs with a stopping-for. In this figure the content creator is explaining a game mechanic, which is when the gryphon, a boss in the game, lands back in the arena, and T3hasiangod tells the players what they should do when this occurs.

[So] the Gryphon comes back down and then you taunt the gryphon and you keep on going against the gryphon so the way that most groups do it is that you attack the gryphon while it's on the ground…

Figure 5.2 T3hasiangod: "So the Gryphon comes back down and then you taunt the gryphon and you keep on going against the gryphon so the way that most groups do it is that you attack the gryphon while it's on the ground…" 6:04.

Deictic gestures can have their own illocutionary force (Ciroux, 2020) as can be seen here when the stopping-for gesture was used for communicative purposes. The stopping of the video is a visual indicator to the audience that the content creator will talk about the still image in front of them. The gesture ties the assertive to the context as the stopping-for purposely focuses the viewer's attention on the gryphon on screen, while the content creator is speaking about it. The deictic gesture itself also takes on a

role, the role of a directive, by conveying the illocutionary force that compels the viewers to look at what is placed or directed-to in front of them. This figure, as well as the others that follow, are deictic gestures that closely resemble directives, regardless of the speech act type that co-occurred in the spoken discourse. The deictic gestures were produced specifically to reference objects, and in a way, they compelled the audience to look at what was being indicated, placed, or emphasized on the screen, much like a spoken command (see also ‘deictic force’ Wu, 2004). This role of deictic gesture use in the YouTube tutorials is similar to its application in the physical world, as exemplified when an individual moves an empty glass at a restaurant to the edge of a table, indicating to the server to fill the glass.

The other most used deictic gesture that co-occurred with assertive speech acts were circular gestures. It is perhaps expected that these would occur the most often with assertive speech acts as well, since they were tied with stopping-for as the most used gesture type. Circular gestures were used to direct the audience’s attention to something specific on the screen. While in other tutorials, this gesture was often used when there were many objects on the screen, in this context, the circular gesture was useful for grabbing the viewers’ attention because it was a constant movement on a paused screen.

Figure 5.3 is an example of an assertive speech act co-occurring with a circular indication. In this figure the content creator is explaining a game mechanic that the gryphon will do to players if they are not careful when it flies into the air.

> They will throw out these little storms here that you can see here so **[these little storms here will]** stun you if they hit you…

Figure 5.3 T3hasiangod: "Now you can see here that the gryphon will fly into the air from time to time… they will throw out these little storms here that you can see here so these little storms here will stun you if they hit you…" 4:53. The black circular gesture line was drawn by me.

T3hasiangod uses a circular gesture on one of the mini yellow circles on the screen when he says, "these little storms here will…" (Image B). This circular gesture was used to indicate which of the yellow circles on the screen is the storm he is referencing (Image A). This deictic gesture was necessary in this situation because the large yellow circle to the left side of the screen in Image A is not a storm that is thrown out by the gryphon, rather, it is a mechanic of the humanoid boss, Siroria. Without the circular gesture, there is a possibility the reference could be mistaken for a different yellow circle as there are several on the screen. This gesture reduces ambiguity in this context and shows the audience exactly which of the yellow circles is the actual storm, thus tying the speech act to the context.

5.2.2 Deligracy's Speech Acts

	Assertives	Expressives	Directives
Circular indication	25 (22%)	1 (5%)	1 (33%)
Pointing	54 (48%)	8 (36%)	0 (0%)
Varied indication	9 (8%)	1 (5%)	2 (67%)
Placing-for	24 (21%)	12 (54%)	0 (0%)
Total	**112**	**22**	**3**

Table 5.4 Deligracy: Counts for the speech acts that co-occurred with deictic gestures. Percentages are based on column-wise totals.

There are three speech acts from Searle's (1976) framework which occurred in Deligracy's tutorial. Assertive speech acts occurred most often in tandem with deictic gestures, with a total of 112 counts. This was followed by expressive speech acts with 22 counts that contributed to the overall total of speech acts that were combined with gestures. Lastly, directives occurred the least with only three counts contributing to the total.

Within assertive speech acts, pointing co-occurred the most with a total of 54 counts, contributing 48% of all deictic gestures that occurred with assertive speech acts. Circular indication occurred in tandem with assertives the second most with 25 counts (22%), followed by placing-for with 24 counts (21%), and varied indication with 9 (8%). For expressive speech acts, placing-for occurred the most with 12 counts, contributing 54% of all deictic gestures that co-occurred with expressives. This was followed by pointing with 8 counts (36%), and circular indication and varied indication each occurring only once (5% respectively). Lastly, within directives, varied indications had

two counts, contributing 67% of the total of deictic gestures that co-occurred with directives. This was followed by circular indication with only one count (33%).

In Deligracy's video, pointing co-occurred with assertive speech acts most frequently. This is unsurprising, as pointing was overall the most used deictic gesture in Deligracy's video, and assertives were the most frequently used speech act. Assertives occurred when the content creator believed that what they were saying is true, or factual. Figure 5.4 is an example of an assertive speech act with a pointing gesture.

[Basically over here we have our aspiration…]

Figure 5.4 Deligracy: "Basically over here we have our aspiration…" 6:05.

This utterance by Deligracy is an assertive speech act because what she is stating is true, the icon she points to is indeed the aspirations tab on the UI. While she is pointing at it, she utters the demonstrative "over here" to verbally indicate to the audience that she is indicating an object, and the pointing visually indicates which object on the UI menu is the one being referenced, as there are many possible choices.

Deligracy was the only one who used expressive speech acts, which are utterances that entail a speaker's feelings towards some entity. These often occurred when Deligracy was giving her opinion about an object on the screen, such as

commenting on houses or whether the wallpaper she was choosing for her in-game house was too bright. Figure 5.5 A is an example of an expressive speech act with a pointing gesture.

[This one] is pretty cute

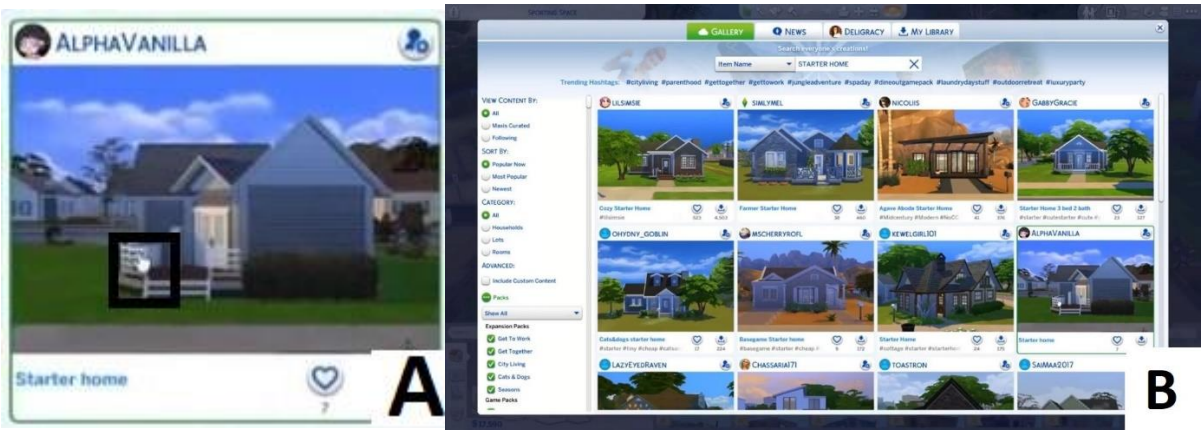

Figure 5.5 Deligracy: “This one is pretty cute” 13.45. The box in A was drawn by me.

In Figure 5.5 B, Deligracy is looking at a menu with many houses that other players have created and uploaded to a network in which players of *The Sims 4* can share their content. Deligracy is looking to pick one of these houses for her sim to buy. In Figure 5.5 A, she directly points to one she is interested in. The pointing co-occurring with the demonstrative “this” gives the expressive speech act meaning, by both verbally and visually indicating to the audience which of the houses in the menu she thinks is cute.

Lastly, directives had only three occurrences and all were suggestions. In these utterances the content creator would advise the audience what she believed they should or should not do, but it was still up to them to decide. For example, in the following directive, the demonstrative “those” co-occurs with a varied indication gesture (Figure 5.6).

they’re just gonna be super unhappy…but I suggest you keep them pretty happy so you know keep an eye on **[those]** make sure they’re good.

This example illustrates how demonstratives and deictic gestures are used to tie speech acts to their context. The demonstrative “those” verbally indicates that a reference is being made and the varied indication gesture visually indicates exactly where on the screen Deligracy is referencing. In this example, Deligracy emphasizes the importance of the needs panel as she had already mentioned it prior to this utterance. The presence of the varied indication gesture is important because without it, the audience would not know what “those” is referencing on the screen (see Figure 5.6), and without the utterance the audience would not know why the content creator is indicating the needs panel.

Figure 5.6 Deligracy: “they’re just gonna be super unhappy…but I suggest you keep them pretty happy so you know keep an eye on those make sure they’re good” 8:10-8:18. The black line and start and end boxes were drawn by me.

5.2.3 BeastyqtSC2's Speech Acts

	Assertives	Expressives	Directives
Circular indication	3 (9%)	0 (0%)	3 (17%)
Pointing	9 (27%)	0 (0%)	4 (22%)
Varied indication	1 (3%)	0 (0%)	2 (11%)
Placing-for	3 (9%)	0 (0%)	4 (22%)
Boxing	17 (52%)	0 (0%)	5 (28%)
Total	**33**	**0**	**18**

Table 5.5 BeastyqtSC2: Counts for the speech acts that co-occurred with deictic gestures. Percentages are based on column-wise totals.

In BeastyqtSC2's gaming tutorial there were a total of 33 counts of assertive speech acts and 18 counts of directives that co-occurred with deictic gestures. Within the assertive speech acts, the gesture boxing occurred the most with 17 counts, or 52% of the total number. Pointing was the next most frequent, with 9 counts, 27% of the total. Circular gestures and placing-for both had 3 counts, contributing 9% each, and varied indication had the least with only one count, contributing 3% to the total number of assertive speech acts that occurred with deictic gestures. Within directives, the most frequent deictic gesture was boxing, with five counts, or 28% of the total. Pointing and placing-for were the next most frequent with four counts apiece, or 22% of the total each. This was followed by circular gesturing with three counts, or 17%, and varied indication which only occurred twice, contributing 11% to the overall total of deictic gestures that co-occurred with directive speech acts. From the data in Table 5.5, boxing was the most frequently used deictic gesture with both assertive and directive speech

acts.

Similar to the other two content creators, the most used deictic gesture in BeastyqtSC2's tutorial from Table 5.1 is also the gesture that most frequently occurs with the most frequently occurring speech acts, assertives. However, in BeastyqtSC2's tutorial, directives (18 counts) occurred more than half as often as did assertives (33 counts).

Figure 5.7 is an example of an assertive speech act co-occurring with a boxing gesture and seven pointing gestures.

"If this **[Pylon dies]**, **[all these gateways become unpowered]**…"

Figure 5.7 BeastyqtSC2: "If this Pylon dies, all these gateways become unpowered..." 21:24. The green box was drawn by me and it outlines the green boxing gesture. The numbers were written by me to indicate the order in which the content creator pointed.

In Figure 5.7, the content creator is explaining a mechanic of the game. The boxing gesture indicates to the audience the ‘pylon’ he is referencing and the following pointing gestures, numbered one to seven in Figure 5.7, co-occur with “all these gateways become unpowered” to indicate which gateways he is referencing that will be affected if the ‘pylon’ his boxing gesture indicated is destroyed. These deictic gestures remove ambiguity by directly pointing out which objects the assertive speech act is referencing.

The main type of directive used by BeastyqtSC2 was suggestion. These were used in BeastyqtSC2’s video when he was advising the audience on what to, or not to do, when building their base. These suggestions are based on his opinions and experience as a seasoned gamer. In Figure 5.8, BeastyqtSC2 is explaining to the audience how to protect themselves from enemy players by blocking an opening on the map near the players’ base. This figure is an example of a directive speech act co-occurring with a boxing gesture.

> The first supply depot that you want to be making as a Terran player is this one **[right here]** and after that you want to follow it up **[with barracks right here]** and then later on whenever you need the second supply depot you **[will make it right here]**, making a full wall-off.

Figure 5.8 BeastyqtSC2: Image A: “the first supply depot that you want to be making as a Terran player is this one right here” Image B: “and after that you want to follow it up with barracks right here” Image C: “and then later on whenever you need the second supply depot you will make it right here, making a full wall-off” 23:11. The underlined phrases coincide with the boxing gesture. The green boxes were drawn by me, and they outline the shape of the box.

In Figure 5.8 BeastyqtSC2 uses “here” three times with three boxing gestures simultaneously. These three sets tie his suggestion to the visual context. The deictic expressions “here” acts as a verbal signal to the audience that a location is being referenced and the boxing gestures not only visually show the audience the locations his suggestion is indicating, they also take on the qualities of a directive speech act, as they draw the audience’s gaze to the location. The boxing gestures are being used to reduce ambiguity and to show spatially where exactly “here” refers to within the visual parameters presented on the screen.

This example also represents how the same deictic gesture that is used to indicate an object on the screen (Image A) can also be used to indicate an object that is not there but will be in the future (Image B and C). The boxing gesture can also be considered to have an iconic component, because the boxing gestures that are used to indicate are also roughly the size of the objects that will be placed in the indicated locations, e.g., the supply depot (Image B) and the barracks (Image C). Since these objects are not there, the content creator is creating a fictive location in the virtual

environment for future action.

All three content creators used assertive speech acts the most and a possible reason for this may be due to the nature of gaming tutorials which inform listeners about some part of a game. Assertive speech acts “express the speaker’s belief” (Huang, 2014, p. 133). In the case of the three gaming tutorials, when explaining mechanics and game structure, the content creators are speaking what they believe is the truth. Expressive speech acts were used in these tutorials when the content creator was expressing their opinion about an object in the game, such as its color or design. This speech act was only present in Deligracy’s tutorial, possibly due to the nature of her game, where one main component of the game is to design, build, and furnish homes.

Directives were used to give the audience suggestions (e.g., Figure 5.6). These were only used by Deligracy and BeastqtSC2 and occurred when the content creators wanted to share their expert opinion about how to play the game or to reach certain goals (e.g., Figure 5.8).

The final aspect that I examined in relationship to speech acts co-occurring with gestures was the length of time that these utterances lasted (see table 5.6). There did not seem to be a distinct pattern as there were times in which the duration of the gesture was shorter in length than the speech with which it co-occurred. An example of this is Figure 5.8, in which the directive speech act was 13 seconds while the series of three boxing gestures was only 6 seconds. Initially I considered the idea that the utterances were long because of post-modification, however, further analysis did not seem to reveal any patterns that suggested this might be the case. Instead, as will be seen in table 5.6 below, the duration of a speech act co-occurring with a gesture was

based on the gesture choice and the explanations of game mechanics or strategies.

5.3 Gesture Analysis

This section discusses select examples that illustrate the different types of deictic gesture observed in the videos and how each content creator uses them. In addition, this section includes which gestures were used most frequently in each of the YouTube gaming tutorials and the possible reasons why some occurred more or less than others with respect to each of the content creators.

5.3.1 Pointing

Even though pointing was the deictic gesture that had the most occurrences in the tutorials, its usage was quite different between the three content creators. T3hasiangod had a much lower number of occurrences of pointing than the other two content creators with a mere two occurrences (see Table 5.1). One possible explanation for such low counts is perhaps due to the nature of the gameplay in this tutorial, which comprises a fast-paced battle scene. Since the visual contents of the video are constantly moving, it is difficult for the creator to use pointing as the objects that he would be pointing at are moving about the screen. Indeed, every instance of a pointing gesture occurred when the video was stopped. This can be seen in Figure 5.9 where T3hasiangod introduces one of the enemies the players must fight but tells the audience that he will come back to them later. The content creator uses a pointing gesture with the mouse while saying:

> **[We'll talk about]** Siroria mechanics later…

Figure 5.9 T3hasiangod: “we'll talk about Siroria mechanics later…” 5:10. The circle was drawn by me.

Earlier I stated that gestures can be used to reduce ambiguity in the visual field when there are many objects that could be the referent for the utterance. However, it is possible that the complexity of the visual field itself could also impact the type of gesture that can be used effectively to indicate. An example of this is with boxing, which has the ability to accommodate how much was happening on BeastyqtSC2’s screen. Boxing seemed to be visually more salient because as a large green box it is easier to see than a small cursor that can be easily missed. Figure 5.10 is an example of a cluttered screen in which the mouse could be easily lost. Contexts such as this may account for why BeastyqtSC2 used more boxing gestures then pointing, which will be discussed later.

Figure 5.10 BeastyqtSC2: Example of mouse not being visible. 21:11.

Deligracy used pointing far more than the other two content creators. A reason for using more pointing than other gestures could be because *The Sims 4* is relatively more UI based than the other games and as a result that is one of the main topics Deligracy discusses. In her video, Deligracy points to the different icons and information on the screen frequently. Many of the objects are immovable, as seen in Figure 5.11, and since the icons are dispersed in different places around the screen, pointing was an ideal way for her to direct her viewers' attention clearly to the different icons on her screen. In Figure 5.11, Deligracy is introducing the different worlds a sim can choose to live in. She first points to Willow Creek and then Oasis Springs.

> **[If you just have the Sims 4 base game you have the option of willow creek] [or oasis springs]**

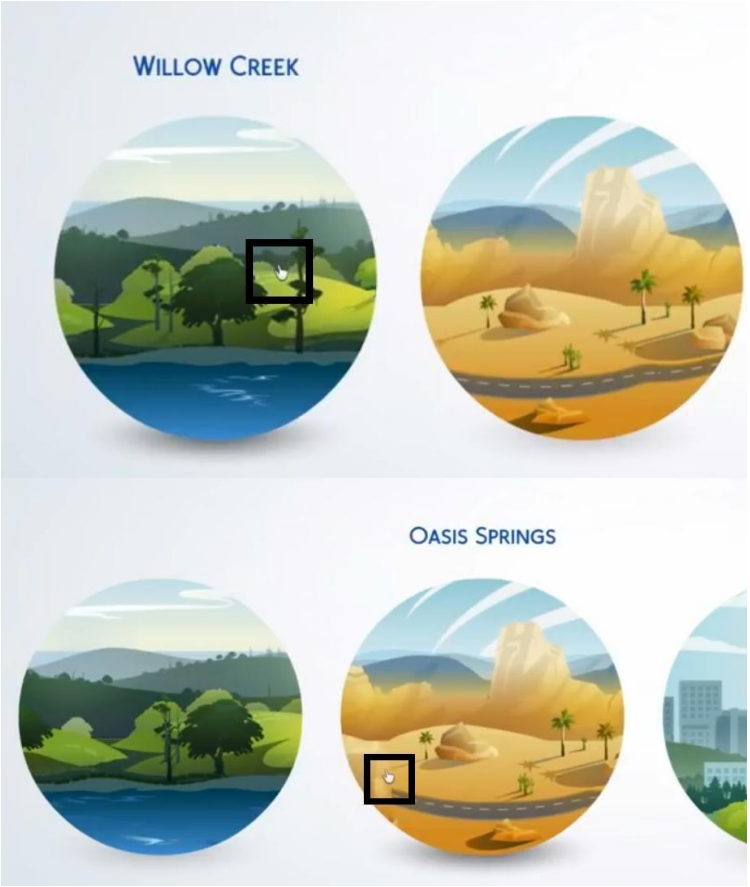

Figure 5.11 Deligracy: “If you just have the Sims 4 base game you have the option of willow creek or oasis springs” 2:07. The boxes were drawn by me.

The explanation of the UI in her game may also relate to why there were a majority of assertive speech acts in Deligracy’s tutorial. When she talks about the UI, she is providing factual information, but some instances of assertive speech acts occur when she gives descriptions. The continuing monologue of Figure 5.11 is an example of this when she describes what Willow Creek and Oasis Springs are:

Willow creek and oasis springs are a desert and more green world

The descriptions she is providing happen to be factual, but they are expressed in her own words and based on her opinion.

5.3.2 Circular Indication

Circular indication in the tutorials was an example of when a gesture is exaggerated to direct audience attention to an object. The content creators put effort into performing the motion of circling, often increasing the effort needed by moving rapidly and repeating the movement several times. Circular indication can make referencing more salient by clearly drawing the audience's attention to a particular place or object, especially in environments that are visually cluttered. By moving the mouse around in a circle, the speakers are emphasizing the location they are indicating. An example of this from T3hasiangod's tutorial is Figure 5.12.

[These little storms here] will stun you

Figure 5.12 T3hasiangod: "These little storms here will stun you" 4:57. The lines were drawn by me.

The circular indication and use of demonstratives tie his assertive speech acts to the context. His utterance is an assertive speech act because he is telling the audience something true about a game mechanic. In this figure, there are many objects on the screen, all of which are immovable. The only way to indicate a specific object amongst the visual clutter is to use some type of directing-to gesture. When he references the specific circle that is on the ground, he uses the proximal “here” as an audio cue accompanied by the circular gesture as the visual cue to direct the viewer’s attention to the correct referent. The object of reference is clear due to the circular gesturing directly pointing out where on the screen the audience should look.

5.3.3 Varied Indication

Although varied indication gestures were used in all three videos, they were never used very frequently, and this was perhaps due to the other available ways of gesturing being more efficient. Partly because of their rare use, they did not seem to have any overarching form or function, therefore they will not be discussed further than what was in the typology chapter (Chapter 4).

5.3.4 Placing-for and Stopping-for

Both placing-for and stopping-for will be discussed together here because they share similar qualities. Since stopping-for was found only in T3hasiangod’s tutorial, it will be explained in concert with his use of placing-for. Due to the nature of how he created his video, which was a pre-recorded gameplay, and his style of providing long spoken explanation, the use of placing-for or stopping-for seemed to be preferred over the use

of many other deictic gestures. The video of the gameplay he discusses is continuously moving, and the only way to talk about what was occurring at specific points in the video at any length was to stop it. In particular instances, additional gestures were used in addition to this stopping-for. In Figure 5.13, T3hasiangod's assertive speech act is used with a stopping-for and circular gesture. He uses both gestures to reference the boss that is on his paused screen. He is telling the audience that the boss will prepare to do another move when they see her raise her sword:

> **[Siroria has another move here]** and that's when she raises her sword into the sky like **[that]**…

Once he stops the screen, he uses a circular gesture with the demonstrative "here" over the image of the enemy, Siroria, to indicate that he will talk about her. He then uses the demonstrative "that" to signal to the audience that what they see on the screen is what he is referencing, and the stopping-for is Siroria raising her sword. What the content creator is claiming is true, the boss will be jumping soon and the pre-action the boss is taking in the game is an indicator of this. To give this speech act context, he uses the demonstrative "here" and "that" to signal to the audience that what they see is what his assertive speech act is referencing, and the circular gesture and the stopping-for provide the visual context. Once again, he uses the deictic behaviors of demonstratives and gestures to tie his assertive speech act to the context.

Figure 5.13 T3hasiangod: "Siroria has another move here and that's when she raises her sword into the sky like that…" 7:43. The black box was drawn by me to show readers the location of the boss in the image.

Placing-for was used when T3hasiangod explained or discussed what was happening on the screen while the video was playing. This was used in situations where it was necessary to show a short duration of what was happening on the screen, when the footage contained what it was that he was referencing. In these instances, the reference is not static on the screen but rather a segment of video, perhaps because the segment was long enough for him to provide an explanation of the referenced object without it being necessary to stop the video. During these segments he did not seem to need to use pointing or other mouse-mediated indications to reference because his use of placing-for was sufficiently unambiguous to be understood. Figure 5.14 is an example of a placing-for co-occurring with an assertive speech act, and it is the continuation of Figure 5.13. T3hasiangod plays the footage of Siroria jumping in Figure 5.14 right after

he finishes his stopping-for in Figure 5.13. In Figure 5.14, the content creator states,

> …she's basically gonna jump up into the air, you **[can see here she jumps into the air…]**

Image A in Figure 5.14 shows Siroria before she is about to jump. Image B shows the same location as A but after she has jumped. T3hasiangod un-pauses the video right after he says "…the air, you…" and while the video of Siroria jumping is playing, he says "can see here she jumps into the air". The use of "here" with the placing-for ties the assertive speech act to the visual context. The content creator is telling the audience that Siroria will jump from their current location to another place on the map. For the utterance to have meaning, it is necessary for the viewer to see this action through the placing-for gesture. The "here" signals to the audience that the placing-for they are watching is the reference.

Figure 5.14 T3hasiangod: "she's basically gonna jump up into the air, you can see here she jumps into the air…" 7:50.

As for Deligracy and BeastqtSC2, both content creators tended to use placing-for

when they wished to reference something that took up an immovable large area. Deligracy often used rotation, zooming, or the placement of an object to refer to what was happening on the screen while BeastyqtSC2 would use an object to reference a certain portion of the screen (see Figure 4.12, 4.13 and 4.14 in Chapter 4 for an example), or the whole screen. In Figure 5.15, Deligracy shows the audience the neighborhood in which her newly created sim will live. The first image is her sim's property where she will build her house. She then zooms out from this location to show us the entire neighborhood, which can be seen in the second image.

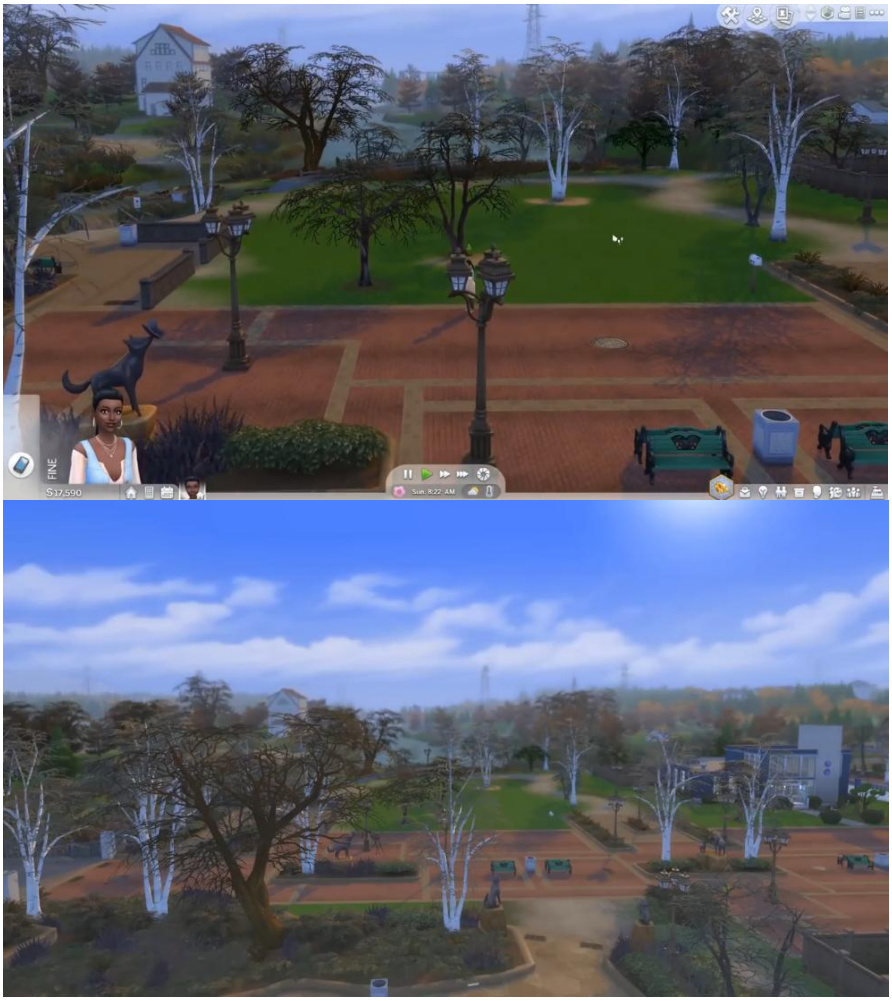

Figure 5.15 Deligracy: "so this is where we're living, isn't it pretty?" 4:57.

The utterance in Figure 5.15 is an assertive speech act that occurs first with a zoom, then a rotation, another zoom, and finally a rotation with a pan of the camera.

[so this is where we're living, isn't it pretty?]

The "this" in the utterance references the neighborhood the content creator has shown us through the placing-for gesture, which lasts for the entire duration of the utterance. The deictic gesture is necessary because without it, the audience would not know what the content creator is referencing. The combination of the gesture and demonstrative tie the speech act to its context by verbally indicating to the audience that what they see in the placing-for of the zoom is what the speech act is referencing.

5.3.5 Boxing

Boxing occurred more than all the other deictic gestures in BeastyqtSC2's tutorial because it seemed to be the manner of indication that helped remove ambiguity the most in this particular game. Boxing could be used for both functional and communicative purposes. It was used functionally as it is a frequent action that is performed to move the game forward, where it serves to select a group of units, a common task in *Starcraft II*. With BeastyqtSC2 being a player himself, he may be accustomed to drawing boxes on the screen, but here, he co-opts these movements in a manner that is communicative. Boxing could also be used to indicate a small or a large area depending on the size and location of the object of reference, providing the user versatility. In addition, boxing had the advantage of being short and quick, much like pointing. Since boxing had many different qualities and potential uses, it is possible

that it was preferred over pointing or circular indication because it was suited more for the type of content that BeastyqtSC2 was demonstrating. The final possible reason boxing may be preferred is that it is visually more salient (e.g., Figure 5.16) than *Starcraft II*'s mouse cursor simply because of its visibility. Boxing is visually separated from the background, thereby helping the audience to draw attention to what matters in context. It is important to emphasize again that boxing is a tool for deixis that is not available for the other games, highlighting how the affordances of particular virtual environments drive specific types of establishing deixis.

Figure 5.16 BeastyqtSC2: "if you have a nexus started… and start warping in units here your unit warping will be much much much faster and it will be a lot easier for you to defend your bases and reinforce as well" 19:33. The black outline around the boxing gesture was drawn by me.

Figure 5.16 is an example of an assertive speech act that occurs with a deictic boxing gesture. BeastyqtSC2 says to the audience,

> if you have a nexus started… and **[start warping in units here]** your unit warping will be much much much faster…

In this example, the content creator tells the audience what he believes to be true, which is if they start warping in units at the location he boxed with his mouse, it should make it easier for the player to defend and reinforce their base. Whether this may be the case or not is unclear until tested, however, from his experience he believes it. The demonstrative “here” and the boxing gesture tie the speech act to the context. The boxing gesture in this example is important because the “here” could be referencing any place seen on the screen. By using this gesture, he is narrowing the visual space to the specific location the demonstrative and speech act are referencing.

All three content creators had similarities and differences in the type of gesture they used. Although they all used pointing, circular indication, varied indication, and placing-for, they each had their own distinct ways of gesturing that was afforded by the games they played and the way they made their tutorials. Boxing was only available in *Starcraft II*, rotation and zooming were only available in *The Sims 4,* and stopping-for was used by T3hasiangod because his pre-recorded video afforded this to him.

Similarly, the three content creators were also alike and yet different in how they used speech acts. Assertive speech acts were the most common across all three content creators. In all three tutorials, it co-occurred with the most used deictic gesture by each content creator. In all three videos, the gestures that co-occurred with demonstratives tied speech acts to their context and gave them meaning. In these utterances, the demonstratives would verbally indicate to the audience that an object was being referenced, and the deictic gestures would visually indicate where or what the object of reference was within the parameters of the screen. This combination of demonstratives and deictic gestures providing clarity by removing ambiguity in YouTube

tutorials is similar to that seen in the physical world (see Kelly, Barr, Church & Lynch, 1999).

An observation from the data showed that visual complexity occurs in these games and the choice of deictic gestures seems to vary accordingly. Some of the factors that added complexity were; the speed of objects moving on the screen (e.g., T3hasiangod's tutorial), the number of objects on a screen (Figure 5.5 B), the visibility of the mouse cursor (Figure 5.10), ambiguity (Figure 5.16), and whether the object of reference was a single object (Figure 5.3), multiple objects (Figure 5.11) or the whole screen (Figure 5.15). The results show an overarching principle, that is, mediated gestures were used to reduce visual complexity and increase clarity in line with the size and nature of the objects being referenced by the content creators' deictic expressions.

5.4 Clark's Temporal Dimension of Directing-to and Placing-for

As mentioned before in Chapter 2, Clark (2003) differentiates placing-for from directing-to using a variety of features that were elucidated in the typology chapter (Chapter 4). Clark also argues that placing-for and directing-to can be differentiated temporally. According to him, placing-for is extended and directing-to is transitory. However, his proposal has not been tested, not in face-to-face contexts nor in virtual environments, therefore, this thesis is the first to test his theory empirically and the results from my research seem to provide support for his claims (Table 5.6).

		T3hasiangod	Deligracy	BeastqtSC2
Circular indication	Range	1.3 - 4.8	0.4 - 4.4	0.6 - 3.4
	Mean	**2.3**	**1.4**	**1.4**
	Standard Deviation	1.2	0.7	0.7
Pointing	Range	0.7 - 1.0	0.1 - 9.9	0.2 - 4.1
	Mean	**0.9**	**2**	**1.3**
	Standard Deviation	0.2	1.5	1
Varied indication	Range	1.5 - 2.8	0.4 - 3.5	0.6 - 2.5
	Mean	**1.9**	**1.5**	**1.6**
	Standard Deviation	0.5	0.8	0.7
Placing-For	Range	1.5 - 6.7	1.1 - 11.2	0.6 - 13.8
	Mean	**3.0**	**3.5**	**4.3**
	Standard Deviation	1.7	2.3	3.7
Stopping-for	Range	1.5 - 3.9	0	0
	Mean	**2.3**	**0**	**0**
	Standard Deviation	0.9	0	0
Boxing	Range	0	0	0.4 - 5.9
	Mean	**0**	**0**	**0.9**
	Standard Deviation	0	0	0.9

Table 5.6: This table shows the duration of the shortest, longest, mean and standard deviation for each deictic gesture per content creator. The mean numbers are bolded to distinctly show the difference in duration for each gesture type. **Duration is in seconds.**

The length of time each content creator held each specific deictic gesture varied. In T3hasiangod's tutorial, the shortest duration was pointing (0.9 seconds), followed by varied indication (1.9 seconds), stopping-for (2.3 seconds), circular indication (2.3 seconds), and with placing-for (3.0 seconds) being the longest. In Deligracy, the deictic gesture with the shortest duration was circular indication (1.4 seconds), followed by varied indication (1.5 seconds), pointing (2 seconds), and placing-for (3.5 seconds). Lastly, in BeastqtSC2's tutorial, the deictic gesture with the shortest duration was boxing (0.9 seconds), followed by pointing (1.3 seconds), circular indication (1.4 seconds), varied indication (1.6 seconds), and placing-for (4.3 seconds).

On average, placing-for took the longest amount of time with T3hasiangod having a mean of 3.0 seconds, Deligracy having a mean of 3.5 seconds, and BeastyqtSC2's having a mean of 4.3 seconds. This acts as support for Clark's theory that placing-for gestures are longer than directing-to gestures. In T3hasiangod's video, stopping-for (2.3 seconds) was on average the second longest deictic gesture. Although it was not the same as placing-for in terms of average duration it still tended to take more time than the other forms of deictic gestures. In Deligracy's case, placing-for was the longest, possibly because referencing an object through rotation and zooming took time. Placing-for in BeastyqtSC2 was similar to the other two content creators in that his placing-for also took the longest. Placing-for and stopping-for were used for explanatory purposes in many instances, and explanations usually take an extended period of time, and this possibly contributed to the duration that these gestures occurred in their video.

Circular indications and varied indications were longer than pointing in T3hasiangod and BeastyqtSC2's videos, possibly due to their form and because of the

way these gestures could be used to emphasize or disambiguate. In theory, both circular indication and varied indications are longer because both require more movement than pointing simply because of the amount of effort used to create their shape. However, circular indication and varied indications were still shorter than placing-for.

The average duration that pointing consumes is shorter than that of placing-for in all three videos. This is parallel to Clark's claim that directing-to gestures are shorter than placing-for gestures. A possible explanation for this is that in both T3hasiangod and BeastyqtSC2's videos, pointing was quick, direct, and straight to the point, therefore making the duration short. Even though this was true in general, in Table 5.6 above, the data seem to indicate that pointing has a longer duration than circular indication and varied indication. This is due to Deligracy having two extraordinarily long points at or around 9.9 seconds as compared to her normal average of 2 seconds. These two points were clearly outliers and skewed the data for the pointing durations. The question then remains of why Deligracy might have performed such elongated points. The physical explanation is that when she pointed to an object, she held her mouse pointer there for a period of time before moving on, and it was this that made the pointing duration longer. A possible reason that she did this could be because she was making use of what Kita (1993) called a *post-stroke hold*, where the indicator is sustained for a period of time thereby prolonging the expression being conveyed by the referencing.

By prolonging the reference, she is letting the viewer know she is still talking about the pointed object. If she were to move her mouse to the next item, the viewers might think she is moving on. In Figure 5.17, Deligracy is introducing the aspirations tab,

located on the UI, which are mini goals the player can try to accomplish to get bonuses for their sims. In Figure 5.17, Deligracy introduces the name of the tab and stays on it for about nine seconds, explaining the sims current aspirations and the rewards the players' sims can attain from finishing these specific tasks.

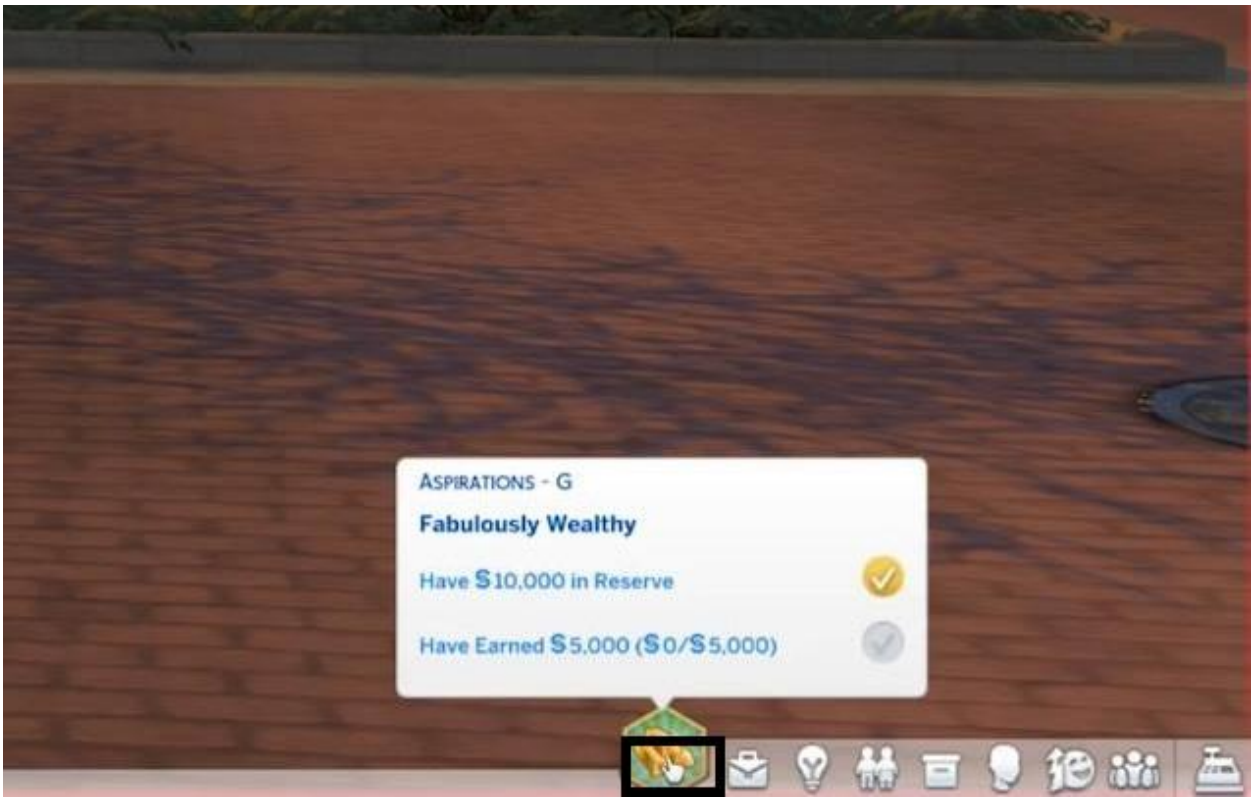

Figure 5.17 Deligracy: "basically over here we have our aspiration and you want to try to fulfill your aspiration so the next thing we need to do is have earned 5000 simoleons…" 6:05. The black box around the mouse location was drawn by me.

Another possible reason for such an elongated point, is that if the content creator were to move her mouse to a different location, the pop-up of the aspirations would disappear making it difficult for her viewers to understand what she is referencing. The pointing in Figure 5.17 is both functional and communicative. It is functional because it causes the pop-up menu to appear, and it is communicative because it visually

indicates what the viewer should look at. This figure is also an excellent example of a double reference where the virtual environment can blend Clark's dichotomy into a seamless whole. She first directly points to the aspiration's icon, thereby drawing attention to that location. She then leaves the mouse in place so that the pop-up menu remains in view for her viewers to reference, which is a classic characteristic of a placing-for. Thus, it could be argued that this single example is both directing-to and placing-for simultaneously.

In other instances, Deligracy would move her cursor to the next object she was going to talk about even if she was still talking about the previous object, as seen in Figure 5.18. This is an example of Clark's (2003) preparatory principle; "most acts of placing-for prepare for the next joint action" (p. 260). What is interesting about this principle is that he discusses its application with placing-for gestures, but in my data, it is also seen with directing-to gestures. By having her mouse hover and stay above the next object of reference, she is letting the audience know that she will be moving on soon, and that she is preparing for the next explanation. In Figure 5.18, Deligracy is describing the jobs that sims can get. Toward the end of her explanation Deligracy moves her mouse to the next object of reference on the UI. The pointing occurs when she states:

We currently don't have a job but we can get a job **[if we want to]**

Figure 5.18 Deligracy: “We currently don’t have a job but we can get a job if we want to” 6:36. The circle was drawn by me.

Lastly, boxing took up the shortest amount of time on average compared to the other deictic gestures in BeastyqtSC2’s video. However, in his video there was an instance of boxing that consisted of an outlier in terms of duration. In fact, it lasted longer than any of the other directing-to gestures and skewed the data for boxing. This particular example of a longer boxing when referencing something could also be considered a form of Kita’s (1993) *post-stroke hold* where the boxing is held to make sure the viewers are looking at the spot he is indicating. This seems to be the strategy that he is employing during the case for his longest boxing (5.934 seconds) in Figure 5.19.

Previously in the video, BeastyqtSC2 talked about game settings and one of them enables text above the structures in the game. In Figure 5.19, his gesture is directing the viewers to notice this setting which tells the player how many workers, or small robots, are currently on the structure. Here he holds the boxing in place while he

is providing an explanation about this feature, and that is why this boxing was longer than the others of this type.

Figure 5.19 BeastyqtSC2: "… is if you have these settings enabled like I was mentioning earlier, you will see how many workers you have right here" 17:19. The large box drawn by me follows the shape of the boxing gesture, and the smaller box outlines the mouse, which turns into the shape of a bracket: [ ].

In general, boxing is a form of directing-to, and this can be seen in Figure 5.19 when BeastyqtSC2 is referencing an immovable object. This particular usage of boxing also shares some similarities with placing-for gestures. The boxing is held for an extended period of time, making it accessible to the viewers. In addition, the continuing presence of the indication makes it clear to the viewers which object is being referenced, while simultaneously acting as a memory aid. All of these features being present at the same time seems to illustrate an instance of a double-reference, where the content-creator uses placing-for and directing-to in the same instance, combining features of each type as some sort of emphasis.

Figure 5.19 is also an example of how deictic gestures and expressions tie speech acts to their context. In this example, the content creator is referring back to something he had mentioned earlier in the video about settings. This situation where the demonstrative “here” co-occurs with boxing is being used in two ways. The first is to reference what is currently on the screen. Both deictic behaviors are important in this utterance because, if either were missing it would be difficult to know what the speech act is referencing as there are many objects and movement happening on the screen. As for the second way the deictic pair were used, earlier in the video he discussed the settings for the game, and one of those settings was to enable the text above the structure. By using the “here” and the boxing gesture, he connects the utterance and reference on the screen to the part in the settings that he talked about before.

5.5 Demonstratives

This section explores the types of demonstratives that co-occurred with each type of deictic gesture across all three content creators. The section also illustrates how many proximals and distals co-occurred with each deictic gesture. Table 5.7 presents the number of demonstratives that occurred with each type of deictic gesture according to content creator.

	Circular indication	Pointing	Varied indication	Placing-for	Stopping-for	Boxing	Total
T3hasiangod	10 / 13 (77%)	1 / 2 (50%)	1 / 5 (20%)	9 / 9 (100%)	8 / 9 (89%)	0 (0%)	**29**
Deligracy	28 / 84 (33%)	59 / 183 (32%)	13 / 30 (43%)	32 / 44 (73%)	0 (0%)	0 (0%)	**132**
BeastqtSC2	7 / 18 (39%)	7 / 32 (22%)	3 / 12 (25%)	13 / 35 (37%)	0 (0%)	19 / 41 (46%)	**49**
	45	**67**	**17**	**54**	**8**	**19**	**210**

Table 5.7: Table shows the total number of counts and the percentage for each gesture type that occurred with demonstratives for all three content creators. [2] Percentages are based on the demonstrative count co-occurring with a specific type of gesture divided by the total number (both demonstrative and nominals) of that specific type of deictic behavior that occurred in the tutorial.

In T3hasiangod's tutorial, 10 out of 13 circular deictic gestures co-occurred with demonstratives, whereas the other three co-occurred with nominals. Placing-for gestures only occurred with demonstratives in nine instances. Eight out of nine stopping-for gestures co-occurred with demonstratives, while only one out of five varied indication gestures co-occurred with demonstratives. As for pointing, only one of the two pointing gestures co-occurred with demonstratives.

Within Deligracy's tutorial, 59 out of 183 pointing gestures co-occurred with demonstratives. 28 out of 84 of circular indications co-occurred with demonstratives, while 32 out of 44 placing-for gestures, co-occurred with demonstratives. Lastly, 13 out of 30 varied indications co-occurred with demonstratives.

In BeastyqtSC2's tutorial, 19 out of 41 boxing gestures co-occurred with demonstratives. Seven out of 32 pointing gestures co-occurred with demonstratives while seven out of 18 circular indications co-occurred with demonstratives. Placing-for

[2] The total counts of deictic gestures in this section do not match that of table 5.1 above because table 5.1 includes all deictic gestures with nominals and demonstratives. However, my focus is on how deictic gestures interact with demonstratives, and the tables in this section reflect that.

had 13 out of 35 tokens occurring with demonstratives while varied indication had only three out of 12 tokens co-occurring with demonstratives.

Table 5.7 above shows how many demonstratives were used with each gesture type, based on content creator. An overarching pattern from the data shows that the most used type of deictic gesture that co-occurred with demonstratives is different for each of the content creators: circular gesturing for T3hasiangod, pointing for Deligracy, and boxing for BeastyqtSC2. This evidences individual differences.

As discussed in section 2.6 in Chapter 2, spatial demonstratives can be broken down into proximal and distal demonstratives.[3] As stated previously, proximals are “this”, “here”, “these” and distals are “that”, “there”, “those”, “over here” (Ensslin, 2012; Huang, 2014; Yule, 1996). The tables below report which deictic gestures co-occurred with the different deictic expressions, and how often. This next section also provides a brief explanation as to the possible reasons why proximal or distal usage occurred more frequently in certain categories of deictic gestures than in others.

[3] I am aware that “this” and “that” do not encode only distance (e.g., Cooperrider, 2016), but I will still use the common terms “proximal” and “distal”, which are standard in this literature.

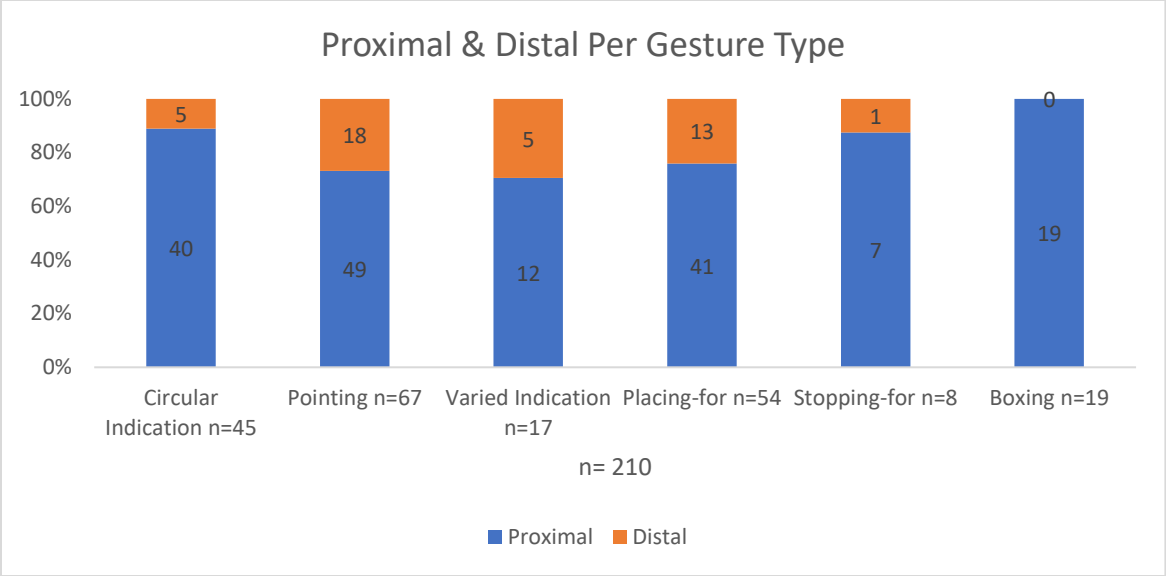

Figure 5.20: Shows the percentages of the proximals and distals together across all three videos.

	Proximal	Distal
Circular indication	40 (24%)	5 (12%)
Pointing	49 (29%)	18 (43%)
Varied indication	12 (7%)	5 (12%)
Placing-for	41 (24%)	13 (31%)
Stopping-for	7 (4%)	1 (2%)
Boxing	19 (11%)	0 (0%)
Total	**168**	**42**

Table 5.8: This table shows the total occurrences and the percentages for the proximal and distal demonstratives that occurred across all three gaming tutorials. Percentages are based on column-wise totals.

Table 5.8 lists the number of proximal and distals that occurred with each type of deictic gesture found across all three videos. There were 210 total demonstratives. Overall, there were more proximal demonstratives (168 tokens) than distal demonstratives (42 tokens). Not only did proximals outnumber distals in general but

they also outnumbered distals in each video (Table 5.9). There are some possible explanations for why this may be the case. One possibility has to do with ambiguity. Recall that Cooperrider (2016) artificially created an experimental context in which ambiguity, which is generally correlated with distance, is decoupled from spatial distance by using a laser pointer. When participants pointed at distant objects with the laser pointer, they nevertheless used the proximal “this” to indicate, thereby suggesting that relative ambiguity or precision is a factor that drives the proximal/distal distinction. Cooperrider’s (2016) laser point is similar to my mouse cursor in that it affords much more precision (to the level of individual pixels) than what would usually be possible with a manual point. Another factor driving the over-representation of proximals may be that the computer screen is actually close from the perspective of the content creator. Third and finally, as seen in Rocca and Wallentin (2020), the use of proximal demonstratives was favored when the object of reference was manipulable. In the case of the games *The Sims 4* and *Starcraft II*, many objects are manipulable because one main aspect of the games is to build and place objects. This aspect of the game, according to Rocca and Wallentin’s (2020) work, could explain the use of proximals by the content creators.

The following section will disaggregate these overall patterns according to the individual content creator’s use.

	T3hasiangod		Deligracy		BeastyqtSC2	
	Proximal	**Distal**	**Proximal**	**Distal**	**Proximal**	**Distal**
Circular indication	10 (37%)	0 (0%)	23 (51%)	5 (11%)	7 (16%)	0 (0%)
Pointing	1 (4%)	0 (0%)	41 (61%)	18 (27%)	7 (10%)	0 (0%)
Varied indication	1 (4%)	0 (0%)	8 (47%)	5 (29%)	3 (18%)	0 (0%)
Placing-for	8 (30%)	1 (50%)	20 (37%)	12 (22%)	13 (24%)	0 (0%)
Stopping-for	7 (26%)	1 (50%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Boxing	0 (0%)	0 (0%)	0 (0%)	0 (0%)	19 (100%)	0 (0%)
Total	**27**	**2**	**92**	**40**	**49**	**0**

Table 5.9: Raw counts and percentages of proximals and distals by gesture type for each of the content creators. Percentages are based on column-wise totals.

In T3hasiangod's video there were a total of 27 proximals compared to two (1%) distals. In Deligracy there were 92 tokens of proximals compared to 40 distals. Finally, in BeastqtSC2 there were 49 tokens of proximals but no distals. As previously mentioned, in general proximals outnumbered distals and this also applied to the individual content creators.

The next table shows the count and percentages of each proximal and distal type that was used for each deictic gesture category across all content creators.

	This	Here	These	That	There	Those	Over here
Circular indication	17 (22%)	14 (22%)	9 (33%)	0 (0%)	1 (25%)	1 (50%)	3 (17%)
Pointing	18 (24%)	17 (26%)	14 (52%)	6 (33%)	1 (25%)	0 (0%)	11 (61%)
Varied indication	6 (8%)	4 (6%)	2 (7%)	2 (11%)	1 (25%)	1 (50%)	1 (6%)
Placing-for	25 (33%)	14 (22%)	2 (7%)	9 (50%)	1 (25%)	0 (0%)	3 (17%)
Stopping-for	3 (4%)	4 (6%)	0 (0%)	1 (5%)	0 (0%)	0 (0%)	0 (0%)
Boxing	7 (9%)	12 (18%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total	**76**	**65**	**27**	**18**	**4**	**2**	**18**

Table 5.10: Demonstrative count and percentage for each deictic expression co-occurring with each deictic gesture. Percentages are based on column-wise totals.

Breaking it down per demonstrative type, “this” occurred the most with 76 tokens. When considering usage of “this” by deictic gesture: “this” co-occurred the most with placing-for with 25 tokens (33%), followed by pointing with 18 tokens (24%), circular indication with 17 tokens (22%), boxing with seven tokens (9%), varied indication with six tokens (8%), and stopping-for with three tokens (4%). The demonstrative “here” occurred the second most with 65 tokens. “Here” co-occurred the most with pointing at 17 tokens (26%) followed by placing-for with 14 tokens (22%), circular indication with 14 tokens (22%), boxing with 12 tokens (18%), and then varied indication and stopping-for with four tokens each (6% respectively). “These” was next with 27 tokens. “These” co-occurred the most with pointing with 14 counts (52%), circular indication with nine tokens (33%), and varied indication and placing-for with two tokens each (7%,

respectively).

As for distals, “that” occurred 18 times. “That” co-occurred with placing-for the most with nine tokens (50%), followed by pointing with six tokens (33%), then varied indication with two tokens (11%), and stopping-for with one token (5%). “There” only had four occurrences. “There” occurred once in each of the following gesture types: circular indication, pointing, varied indication, and placing-for and each accounted for 25% of the total. “Those” only had two instances, with one token in circular indication and one token in varied indication, each making up 50% of the total. Finally, “over here” occurred 18 times. “Over here” co-occurred with pointing 11 times (61%), circular indication and placing-for both had three tokens each (17% respectively), and varied indication had only one token (6%).

The previous section displayed the occurrence of proximals and distals across all gesture types. This section, and Table 5.11, will present the most frequently used proximal and distal demonstratives by each individual and explore the possible reasons for the usage.

Content creator	**Proximal**			**Distal**				**Total**
	This	**Here**	**These**	**That**	**There**	**Those**	**Over Here**	
T3hasiangod	7 (24%)	19 (66%)	1 (3%)	2 (7%)	0 (0%)	0 (0%)	0 (0%)	**29**
Deligracy	41 (31%)	28 (21%)	23 (17%)	16 (12%)	4 (3%)	2 (2%)	18 (14%)	**132**
BeastyqtSC2	28 (57%)	18 (37%)	3 (6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	**49**

Table 5.11: Proximal and distal with all deictic gestures according to the content creator. Percentages are based on row-wise totals.

Looking at proximals first, the type of proximal T3hasiangod used most was “here” with 19 occurrences, or 66% of all demonstratives used in his video that co-occurred with deictic gestures. This is followed by seven tokens (24%) of “this” and one token (3%) of “these”. “This” occurred most frequently in Deligracy’s video, with a total of 41 tokens, or 31% of all demonstratives used in her video that co-occurred with deictic gestures, followed by “here” with 28 tokens (21%) and 23 tokens of “these” or 17%. Lastly, BeastyqtSC2 had the most occurrences of “this” with 28 tokens, or 57% of all demonstratives that co-occurred with deictic gestures in his video. This was followed by “here” at 18 tokens (37%), and “these” with three tokens or 6% of all demonstratives.

The majority of distals were produced by Deligracy. T3hasiangod only used “that” twice, making up 7% of total demonstrative used that co-occurred with gestures in his video, while BeastyqtSC2 did not use any distals. Deligracy, on the other hand, had the most use of distals with 40 tokens making up 31% of all distals that co-occurred with deictic gestures in her tutorial. In her video, “over here” occurred 18 times (14%), followed by “that” occurring 16 times (12%), “there” occurring four (3%) times and “those” occurring two times (2%).

Each content creator has a different number of tokens and a different pattern of usage. Perhaps the most interesting pattern that emerges from the data above is the clear division between the use of proximals and distals. Distals occur much less frequently than proximals and nearly all of them are produced by a single content creator. There were a total of 42 distals (18.8%) that co-occurred with deictic gestures found in the data and all but two of them occurred within Deligracy’s video. The interesting question, therefore, is what made her video so different from the other two.

Even though there is no traditional sense of distance in *The Sims 4*, as seen in the physical world, there is still a sense of relative distance and temporal relative distance within the game itself. Table 5.12 shows how many distals were used in each of those categories within Deligracy's video.

	Relative Distance	**Temporal Relative Distance**
Distal	24 (60%)	16 (40%)
Total	**40** **(100%)**	

Table 5.12: shows the two types of distance found in Deligracy's tutorial on *The Sims 4.* The use of distals were based on these two contexts. Percentages are based on column-wise totals.

There were two types of relative distance found within the tutorial. The first type of in-game relative distance was the indicating of objects from the content creator's perceived distance between themselves, in the physical world, and the object in the game. The second type of relative distance was the indicating of an object from the first object's origo in the game to the second object in the game. This could be from the origo of the sim, the perceived position on the screen indexed from the mouse or the first object that was pointed to initially by the content creator. A total of 24 (60%) of the distal tokens that co-occurred with gestures seem to have been used because of these two types of relative distance. The type of distal that Deligracy used the most was "over here" with 18 tokens. Aside from its use in relative distance, "over here" often appears to be used when there are many different objects on the screen, and Deligracy is letting the viewer know that they should shift attention to the new place she is going to

reference.

Figure 5.21 is an example of this when Deligracy is introducing and describing the different worlds, or neighborhoods, her sim can live in. This example follows the events of Figure 5.11. In this portion of the video, she is talking about Willow Creek, Oasis Springs, and Newcrest. She first only mentions Willow Creek and Oasis Springs by name, and this was seen in Figure 5.11. She then introduces Newcrest. After introducing Newcrest and giving a description about it, she goes back to describe Willow Creek and Oasis Springs as she had not done so when she originally introduced them. Once the description for them is over she points to another world with her mouse called Magnolia Promenade. The use of the pointing gesture on Magnolia Promenade co-occurring with the distal “over here” signaled to the audience that she is moving on and that she is indicating both verbally and gesturally where the audience should shift their attention.

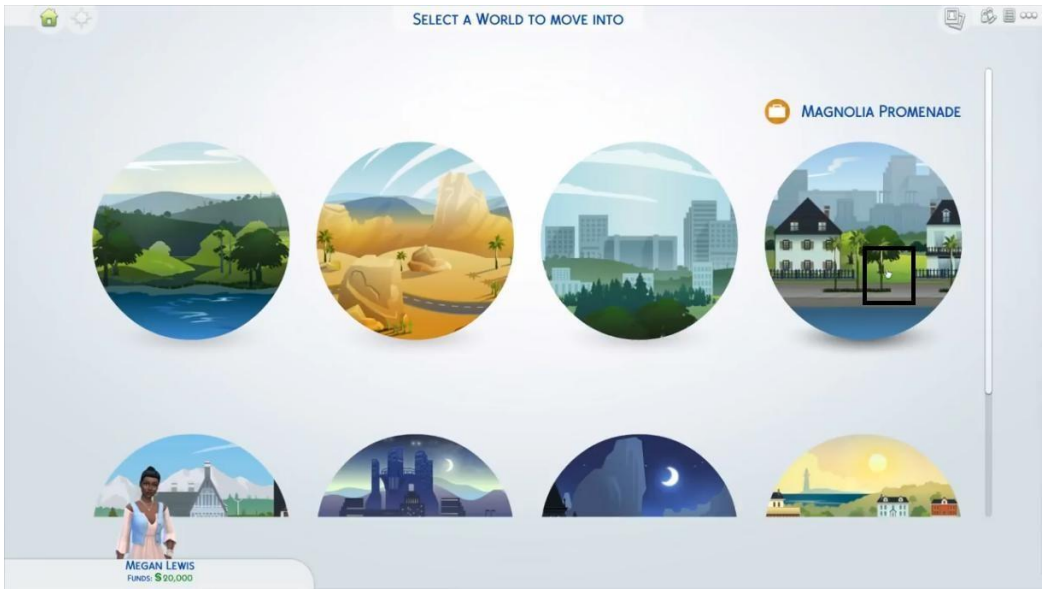

Figure 5.21 Deligracy: “Over here we have magnolia promenade…” 2:24. Deligracy points to the Magnolia Promenade world icon with her mouse as she says “over here”. The black box around the mouse was drawn by me. Figure 5.11 earlier in this chapter is the prequel to this.

A further example (Figure 5.22) that uses the demonstrative “over here” illustrates how relative distance and deictic shifting interact.

If we go **[click over here]** she’ll actually walk over there

In this instance, the content creator is an example of being doubly-situated (Ensslin, 2009). She first indicates a location on the screen from her perspective but then changes to indicating the same location from her sim’s position, possibly showing a temporary immersion, or spatial deictic shift, into the game world. The first “over here” co-occurring with a pointing gesture is indicating a location on the screen from the origo, or perspective, of the content creator. The content creator then deictically shifts from her origo to her sim’s when she uses the distal demonstrative “over there”. The distal is used to indicate the relative distance from the sim to the indicated location. This is evident as the location of the “over there” is the same location in the virtual environment as the “over here”. The content creator could have used “here” again but she does not because she has shifted the origo. In addition, the deictic shift is further evidenced by how, in the first indication, the content creator uses the inclusive first-person plural pronoun “we” that encompasses the content creator and the audience, as they are the only ones who can click on the screen. However, in the utterance where she says “she’ll”, the content creator uses the singular pronoun representing the sim. This change in pronoun use also substantiates the argument that there was a change of perspective.

Figure 5.22 Deligracy: “if we go click over here, she’ll actually walk over there” 9:06. Black box was drawn by me.

Figure 5.23 is another, albeit different, example of deictic shifting seen in Deligracy’s tutorial.

Figure 5.23 Deligracy: “thank you so much for the fruit cake, that’s nice you guys” 12:00. Black box was drawn by me.

In this example, the content creator is referring to and shifting themselves between both the physical world and virtual environment. The content creator says, “thank you so much for the fruit cake, that’s nice you guys” to two of the in-game characters. The content creator is engaged in the narrative and she slips into the virtual environment and thanks them as if she was her sim, who was the one that actually received the fruit cake. This type of deictic shifting could be considered doubly-deictic (Bell et al., 2018; Herman, 2002) in which the content creator is speaking from both her physical self and from her sim’s perspective. The sim is the one who received the fruit cake and is in the same world as the other characters, but the content creator identifies with her sim and feels glad to have received it and so shows appreciation as if the game characters had given it to her instead.

Figure 5.23 is also an example of an expressive speech act that is tied to its context by the deictic expression and gesture. The content creator points to the fruit cake with her mouse while she expresses her thanks. The pointing visually indicates what her speech is referencing, which also ties the doubly-situatedness and the deictic shifting of perspective to the context, as they are co-occurring when the content creator is referencing the fruit cake.

One focus of DST and doubly-situated theory has primarily been concerned with narrative, either from text, such as poems (AbdulilahGheni, 2019), or games with a built in narrative, e.g., digital fiction (Bell et al., 2018). However, *The Sims 4* tutorial, which was the only video that had deictic shifting, does not have a built-in or programmed narrative. This shows that DST can be applied and is present in contexts, e.g., YouTube gaming tutorials, that do not have nor need a built-in narrative.

The other type of distal that Deligracy makes use of is a type of temporal relative distance. The action of building and placing objects in *The Sims 4* is divided into at least three phases. One phase is when the player selects the object they are going to build or place, another is when they decide where to place the object, and the last phase type is when they decide on the design of the object.

When Deligracy is engaged in the act of manipulating something within the game, whether the phase is continuous or discrete seems to determine which type of demonstrative she will use. The first two phases are the selection of the object and the placement of the object, which can take a longer amount of time because the player is trying out different positions and orientations of different objects while trying to make their decision. These two phases are continuous phases in that the object must move from point A, such as the UI bar, to point B, a location in the house, by traversing the space in between, much like a click and drag across the screen. When Deligracy is engaged in the act of manipulating something within the game, if the phase is continuous, she tended to refer to the object with proximal demonstratives because the action was still ongoing (Figure 5.24 A). However, in the second phase, once the action was no longer continuing, e.g., the object had been positioned, if she referenced that object again, she tended to refer to it with a distal demonstrative (Figure 5.24 B).

When Deligracy is choosing the types of chairs she would like, and has selected them from the menu, she uses the proximal “these” with the continuous action of placing them. Once she finishes the action of putting the chairs next to the table, she uses the distal “that” to reference the action of placing the chairs.

Figure 5.24 A Deligracy: "…lets get some chairs, these are kinda, kinda funky little chairs lets just go like that" 21:51. This scene exemplifies the continuous phase where she is selecting and placing the object and referring to it with the proximal "these". When she has completed the phase of selecting and placing the object she then refers to it with a distal "that".

The only changes that are left to make after the object is positioned, are discrete changes that are instantaneously completed by the game software itself and this lends itself to distal demonstrative use because the action is completed. For example, in Figure 5.24 B, when the content creator is changing the colors of the chairs she just placed, the choosing of colors on the color palette does not need to fade through all the intervening shades between color A and B; it simply changes from color A to color B. This means that once a color has been pointed to, that action is complete, even if the content creator points to another color later, each point is a completed action.

Figure 5.24 B Deligracy: “do we wanna do that do we wanna do that do we wanna do that do we wanna do that oh my gosh so many choices” 22:00-22:05. In these figures she is trying to decide what to change the chair color to. This scene exemplifies the discrete phase.

Figures 5.24 A and B demonstrate that the choice of distal over proximal may be chosen based on whether an action the content creator is performing is continuous or discrete. When Deligracy is choosing an object and is in the transition of placing it in the first phase, she tended to use proximal. However, once the placement was completed, in the second phase, she tended to use distal to indicate the same object. One similarity that phase two and phase three had was the completion of an action, and in phase three when discrete actions were completed, the content creator tended to use distal to reference them.

The only other instances of distal usage occurred in T3hasiangod’s video, and they only occurred twice, once with a placing-for and once with a stopping-for. One possibility for why he uses the distal “that” instead of a proximal could be because he habitually used the proximal “here” to refer to everything that is happening on the screen in general. There are also other possible factors that may have influenced his use of distals in these two instances.

In Piwek et al. (2008) distals occurred with objects of low deixis, which are referents that are not difficult or require much effort to find. He also showed that distals were “preferred to refer to entities with high accessibility” (p. 702). These two possible reasons for distal use seen in Piwek et al. (2008) may explain the use of distal for the two instances used by T3hasiangod. The object of reference of placing-for and stopping-for are seen on the screen itself. This means that there is not always a need for a directing-to gesture. This also makes the placing-for and stopping-for highly accessible, and because of this, the audience does not need to put in effort to look for the referent because it is right in front of them.

Unlike the other two, BeastyqtSC2 did not use any distals in his tutorial video. It is not immediately obvious from the game mechanics why this would be the case, as the game he plays does not differ with respect to relative distance within the game. It could perhaps be an individual difference or an idiosyncratic preference by this content creator.

5.6 Demonstrative Choices

In the physical world, objects that are at a distance tend to be ambiguous when being pointed to with a hand because the speaker has to project a vector from an indicator, such as the tip of the finger, to the object of reference (Hassemer & McCleary, 2018; Kita, 2003). The farther the distance, the wider the vector becomes and possibly because of this more objects in the distance are covered by the vector making it difficult to know which should be taken as the object of reference, thus creating ambiguity. In contrast, pointing to objects that are closer to the speaker tends to be less ambiguous

because the vector is smaller and there is not as much, or even anything, to interfere with the indication. In the YouTube gaming tutorials, however, a vector is absent and everything can be pointed to, thus there is no far distance, yet the creators still made use of distal demonstratives as can be seen above. This seems to support my argument that distals are not only used in relation to distance, and that distance is not the only factor that affects demonstrative choice, in fact, I argue that ambiguity plays a significant role. Even though everything on the screen is close, and can be pointed to with the mouse, it does not mean that there is no ambiguity and this plays a role in how the content creators chose to use demonstratives and deictic gestures in their videos.

As seen in Table 5.8, proximals were used mostly with pointing. A possible reason for this is the precision and clarity afforded to pointing with the mouse. An example of its affordance can be seen in Figure 5.25 where the content creator is using a pointing gesture to indicate the start game icon on the UI, and she places the mouse pointer on top of the object.

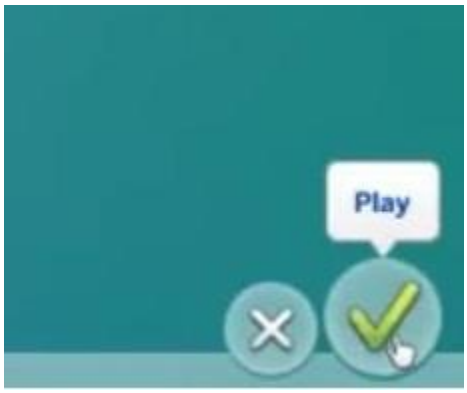

Figure 5.25 Deligracy: “you're just gonna go over here to the play button" 1:03.

The mouse pointer is an extension of the content creator in the game, and as seen in Figure 5.25, there is no far distance between the pointer and the objects in the

game as there is no vector coming from it. The pointing in this example is precise, making it unambiguous because the mouse pointer is right on top of the object.

Although a pointing with the mouse is often precise and clear, there are some circumstances in which the mouse pointer can become too precise, and this could lead to the content creators using other forms of indication, such as circular indication or boxing. An example of this is seen in BeastyqtSC2's game where the visuals on the screen were more crowded. Given that it is a tiny object on a big screen that is quite visually cluttered, the cursor becomes lost and is not easily distinguished from the other objects on the screen and thus is unable to perform its function of directing attention properly. This weakness of pointing is one of the possible reasons why all three YouTube videos had circular gesturing and why there was more boxing with proximals in *Starcraft II* than any other type of gesture. An example of this is in Figure 5.26 in which BeastyqtSC2 is suggesting to the audience where they should build a spore crawler in order to defend against enemy units.

> If you have to make a spore crawler against enemy flying units, make it about **[here]**, **[right here]**….

The content creator first uses a pointing gesture, with the demonstrative "here" to indicate a location on his screen. He then quickly uses a boxing gesture to reference it again while saying "right here". His initial pointing gesture co-occurring with "here" can be hard to see because there is so much happening in that location. It is possible that BeastyqtSC2 realizes this ambiguity and so he quickly gestures again with a boxing and another demonstrative to remove the ambiguity. The way boxing was used with the

proximal “here” in Figure 5.26 ties his speech act, a suggestion, to the context. The first “here” verbally signals to the audience that he is indicating something on the screen. The second “here” is used to re-indicate the location. With the use of the term “right” before “here", he is emphasizing that location to provide clarity. To make it easier for the audience to see where his “here” is indicating, he incorporates a boxing gesture.

*Note the difference in size of the gesture when comparing the first set of figures (Figure 5.26 A - pointing) and the second set of figures (Figure 5.26 B - boxing). The boxing covers a larger area and encompasses more than just the pointing.

Figure 5.26 A BeastyqtSC2: “If you have to make a spore crawler against enemy flying units, make it about here, right here...” 29:20. In this set of figures, the “here” is accompanied by a pointing gesture. The top figure is the original size (100% in ELAN) of the screen. The bottom figure is a zoomed in version (200% in ELAN). The arrows, circle and boxes were drawn by me.

Figure 5.26 B BeastyqtSC2: In this set of figures, the “right here” is accompanied with boxing. The top figure is the original size (100% in ELAN) of the screen. The bottom figure is a zoomed in version (200% in ELAN). The arrows, circle and boxes were drawn by me. The box shape drawn by me is the outline of the box shape created by the content creator in the game.

When BeastyqtSC2 uses boxing, the audience can clearly see the area he is referencing because the boxing highlights the entire area with a clearly visible green box. This was not necessarily to reference the entire area of the box, but rather to simply isolate the part of the screen where he wished to indicate something specific. By using boxing, he decreased the amount of area that the viewer needed to scan in order to find the specific location that he was referencing. This usage of boxing could be considered vague because of how much area it covers as well as how it is not a direct point, however, it makes referencing an object more noticeable in a cluttered game world because of how it pulls the audience's view to the area containing the referred

object, especially since the box highlights the selection in another color. In cases such as Figure 5.26 B, boxing makes referencing unambiguous, and this possibly accounts for why boxing always occurred with proximals.

Figure 5.26 is another example of how deictic expressions and gestures tie speech acts to their context. Both sets of demonstratives and their corresponding gestures, the first “here”, co-occurring with a pointing gesture, and the second “here” co-occurring with the boxing, gives the speech act meaning, but in different ways. The first set signals to the audience that his suggestion is referring to a location on the screen, however, as it may either be difficult to see, or for emphasizing purposes, he adds a second set of indication to clarify. Without the second set of indication, the audience might not know where he was referencing, or the audience may think the initial “here” is referencing the whole screen as a placing-for since the original mouse point was not very visible.

In addition to directing-to gestures influencing proximal choices, placing-for and stopping-for gestures that occurred with proximals were precise and clear in their own ways. Instead of having a single point on the object, the placing-for and stopping-for were used to reference the entire screen, something within the screen that the creator assumed the viewers already knew about, or a video demonstration related to the object of reference.

In Figure 5.27, it is clear what the object of reference is because the content creator uses the placing-for to create joint accessibility where anyone watching the video can see the referred object. It is further clarified when he says, “you just lift your barracks like this…” and then he proceeds to show the barrack move from one position

to another on the screen.

Figure 5.27 BeastyqtSC2: “you just lift your barracks like this…” 25:18. BeastyqtSC2 plays the tutorial to show the audience the movement of the barracks. The arrow was placed by me.

The placing-for here is also an important component to understanding the content creator’s speech act. In this scene he is suggesting to the audience where they should put the starting barracks, and that later in the game they can move it (as seen in the figure). To show this, the content creator gives a visual demonstration, which is the placing-for. “This” is used to verbally indicate that the movement the audience sees on the screen is what the content creator’s suggestion is referencing. The visual effects of the placing-for is a demonstration for the audience of what moving the barracks visually looks like. The use of “this” with a placing-for signals to the viewers that the barracks and the movement of the barracks on the screen is the object of reference, and this combination of deictic behaviors give the speech act substance.

In the case of stopping-for, pausing the screen was a way to say, ‘this scene I just stopped on is what I will talk about’. The reasons proximal was used with placing-for can also be applied to stopping-for. It was assumed that the viewers could infer that when the video stopped, the content creator was going to talk about something in the

footage, making the stopping-for unambiguous. The screen was in the direct view of the audience, and due to the context of the explanation and what was being said there was no ambiguity as to what was being referenced and this seems to be why a proximal was used. If the video continues playing, there would be more ambiguity as to which of the activities happening on the screen was the object of reference. In addition, this type of gesturing is quite a different form of reference in that the entire scene is referenced, and the audience is invited to look at it and find out what is relevant. This is different from the other cases, e.g., pointing, where a particular object within a scene is singled out.

Stopping-for deictic gestures were often used when there was a forthcoming long explanation or when T3hasiangod wanted to talk about actions that were happening on the screen. Figure 5.28 is an example where T3hasiangod pauses the screen to create a stopping-for. The stopping-for and the use of the proximal “this” in his sentence tells the viewer that what they are looking at is the boss named Siroria. He does not need to point to anything because they can see the boss in the center of the screen. Though there are other player avatars visible on the screen, it seems that the content creator is assuming those who are watching know that the blue bars over those other characters indicate that they are actual players on the same team and that the boss is the one in the distance. Although there is distance in the sense that Siroria does appear to be far from the players, the creator still uses proximals because there is no ambiguity in this instance. He expects the people watching the video to have a shared common ground and know that Siroria can only refer to the non-player character on the screen.

[This is one of the side bosses this is Siroria]

Figure 5.28 T3hasiangod: “This is one of the side bosses this is Siroria” 1:18. The circle and arrow were drawn by me.

Lastly, circular indication was used in a distinct way, not simply in instances where there was clutter in the virtual environment. Rather, it was used to reference a general category of objects, a type of metonymy in which the specific stands for general. An example of this is seen in Figure 5.29 A in which BeastyqtSC2’s first circle, co-occurring with “these”, goes around one of the barracks that is currently on the screen blocking off a path. In this example, he seems to be using the proximal “these” to reference a general category of all of these types of barracks, not specifically the single barracks that he points to. This is similar to when someone is explaining different balls in a sports game, and they give a demonstration with only one ball in their hand while saying for example, “these kinds of balls need to be handled with care”. The “these” is a single example, but it is instantiating a general type, which may explain why the plural is used.

[So these starting barracks…]

Figure 5.29 A BeastyqtSC2: “So these starting barracks…” 25:13. The lines were drawn by me.

[…should always go like this]

Figure 5.29 B BeastyqtSC2: “...should always go like this…” 25:15. The lines were drawn by me.

In Figure 5.29 B, BeastyqtSC2 uses the proximal “this” with a circular gesture on a general area that is next to the barracks he had just referenced in Figure 5.29 A. The “this” here is not being used to reference the object but rather in conjunction with the circular gesturing it is indicating an area on the screen that he is trying to draw the

players' attention to. In a sense he is telling the viewers to look at what he is pointing to, and not anywhere else. Like the previous examples above, the use of the demonstrative and deictic gesture ties the content creator's speech act, in this case a suggestion, to the context. The first circular gesture that co-occurs with "these" is indicating a specific object, the barrack. The second circular gesture that co-occurs with "this" is not only indicating a location for the audience to look at, but the words he uses, "should always go like this" informs the audience that what they see on the screen is an example of how barracks should be placed to wall off an opened area.

5.7 Deictic Gestures as Communicative Devices

Not all body movements are considered to be gesture that is genuinely communicative movements that are part of utterances (Kendon, 2004). Instead, many body movements are purely functional and goal-oriented, such as actions performed on objects that are not interpreted communicatively (Novack et al., 2016). The same applies to mouse movements and the placement of the objects in YouTube gaming tutorials, which may be purely functional and goal-oriented, such as fulfilling a purpose in the game, or in comparison to uses that are intended to be communicative.

The mouse is a functional device that moves things forward, and it is also a communicative device used to reference an object, much like pointing in the physical world. The question is how did the content creators differentiate between functional and communicative uses of the mouse? An example of functional use is in Deligracy and BeastyqtSC2's games where the mouse was often used as a way for the players to move the game forward toward a goal. For example, in Deligracy's video the goal of the

game section was to build and furnish a house for her sim to live in. During this section Deligracy picked up and moved objects around or placed them into the game world. By doing this she moves the game along for her tutorial so she can move on to the next part of her discussion. The functional movements she makes are what Novack et al. (2016) call *goal-oriented gestures.* Figure 5.30 is an example of this when she picks a curtain from the furniture menu and places it onto the window in her house.

Figure 5.30 Deligracy: “maybe we can just like drape a nice curtain just like that.” 22:25.

There were a few ways by which the content creators signified that a mouse movement was communicative rather than functional. One factor in which a mouse cursor movement may be interpreted as communicative is via the linguistic context. If a speaker uses a deictic term such as “this” or “over here”, these terms are an auditory

signal to look at what they are indicating. These terms are regularly used together with pointing in the physical world (Cooperrider, 2016), and so viewers might expect a deictic gesture when these words are used.

Novack et al. (2016) observed that a movement of the hand that had no functional purpose could be considered communicative if the “movements can be interpreted as meaningful” (p. 345). This is also evidenced in the YouTube gaming tutorials, such as circular indications and stopping-for, which were communicative, and had no functional purposes. It was clear these gestures were used for communication purposes simply because using these actions accomplished nothing in the game. In the case of circular indication, this gesture took the form of circular rotations around an object of reference, so that the object could be seen by the audience. It could be argued that the exaggeration of the circular motion was a way to communicate with the audience. In the physical environment, individuals exaggerate a movement to make it clear that the viewer should pay attention to it, which invites the inference that the gesture was meant to be seen, and thereby communicative (Fauville, Luo, Queiroz, Bailenson & Hancock, 2021). In the gaming tutorials, the content creators exaggerate mouse movements to emphasize and reference a location or object on the screen, thus making it a communicative deictic gesture.

In the case of stopping-for, it was evident that it was used for communicative purposes simply because when T3hasiangod stops the video, he talks about what is on the screen, and when he finishes his discussion, he starts the video again. Since his tutorial was not a live gameplay, whenever he references something on this screen, either through a directing-to or placing-for, he is communicating with the audience.

None of the mouse movements he made are functional to move anything forward in the game. Instead, the only functional movements for his tutorial were the manipulation of Windows Media Player, such as playing, stopping, or fast forwarding the video.

Boxing is an example of a mouse movement that had both functional and communicative purposes. It serves a purpose in the game, that of selecting multiple units (robots) quickly in one movement of the mouse. However, BeastyqtSC2 used boxing in a new way to communicate with his audience. He used it to reference and redirect the viewers' attention by boxing locations, or objects without necessarily performing further actions with them. Another example where a deictic gesture was used both functionally and communicatively was found in Deligracy's tutorial. There were times where she referred to an object with the mouse while moving the game forward (see Figure 5.31 A & 5.31 B). In these examples, Deligracy states that she is going to put a dining table in the corner. She indicates to the table and chairs with her mouse pointer (communicative device), while at the same time placing them down (functional device).

Figure 5.31 A Deligracy: “let's just get this like, little basic white one” 21:48. She places the table down.

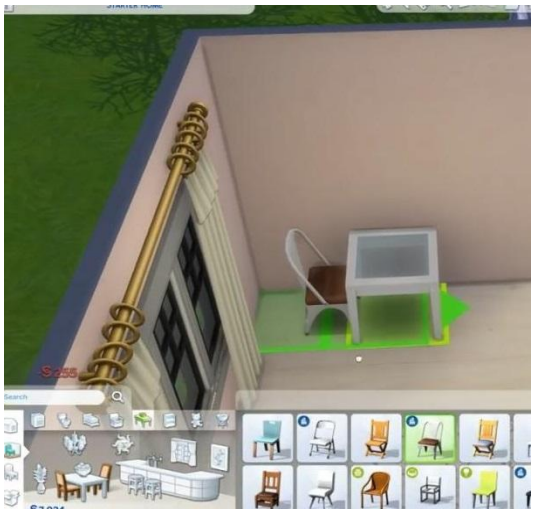

Figure 5.31 B Deligracy: “lets get some chairs these are kinda kinda funky little chairs” 21:50. She places the chairs down.

In this chapter, I have made the case that much of the distinction between placing-for and directing-to can be carried over to the virtual environment, but the distinction also breaks down in this space. I found that it is not always possible to separate the poles of the dichotomy from each other as the virtual environment allows instances which contain elements of both (e.g., Figure 5.17). The findings in my data also seems to show that in the virtual environment the mouse is similar to the function of the laser pointer in Cooperrider (2016), in that gesturing with the mouse also removes

ambiguity, which influences the choice between proximal and distal demonstratives that co-occur with gestures. The results from data found in the YouTube gaming tutorials help support Cooperrider's (2016) claim that distance and ambiguity should be decoupled from each other. I believe this was accomplished through my presentation of a repertoire of other types of deictic gestures that remove ambiguity. If one gesture was not applicable to a situation to remove ambiguity, such as a cluttered environment, the content creators used other gestures such as placing the object in front of the audience to see or emphasize a location by moving the mouse in a way that will attract the audience's attention. These types of deictic gestures contribute in their own way to remove ambiguity.

The use of proximals outnumbered distals for each content creator, and although I argue that proximal choice was mostly due to the lack of ambiguity in the gaming tutorials, there were other possibilities for their increased use, such as the manipulability (Rocca & Wallentin, 2020) and the placement (Coventry, Valdes, Castillo & Guijarro-Fuentes, 2008) of objects. The use of proximals mentioned by both studies applied to the data set in this thesis, as the objects the content creators are manipulating and placing are the objects they are discussing and referencing, which can be visibly seen by the audience (e.g., Figure 5.27). However, I expand on the previous research by discussing what happens after the object has been manipulated and placed. The findings showed that distals tended to be preferred when an action was automatically completed, or when content creators re-referenced a placed object (e.g., Figure 5.24).

Although distance, in the traditional sense, did not necessarily exist in the tutorials, I found there were still forms of distance within the games, and these were

what I called relative distance and temporal relative distance. Proximal choice in the tutorials was based on relative distance such as that found in the perceived distance between an origo and the object of reference, or it was based on an action in progress. In contrast, once these actions were completed and the object was to be referenced again, distals would sometimes be used. This contributes to the literature on demonstrative use, as to my knowledge, distals have not been discussed as being markers for re-referencing completed actions. These observations on proximal and distal usage show that the traditional idea that proximal is used to reference close objects and distals are used to reference far objects is not all-inclusive (Coventry, Griffiths & Hamilton, 2014), and this opens up doors for further research into the use of demonstratives in the virtual environment.

Previous research investigated the use of temporal and personal deixis co-occurring with gestures in pragmatics or with speech acts. However, there has been little research on the relationship between spatial deixis, deictic gestures, and speech acts. The investigation on speech acts in this thesis showed that spatially deictic expressions do occur within speech acts and along with deictic gestures, give meaning to the utterances.

A final contribution of this thesis was to empirically demonstrate how Clark's (2003) placing-for and directing-to dichotomy are temporally distinct from each other. Using actual duration measurements from my data I was able to demonstrate that placing-for is a continuous gesture type while directing-to is transitory, at least in the virtual environment. By extension, it is possible that the data here can inform future work.

Chapter 6: Conclusion

As humans, we use whatever means are present to establish communication in whatever environment we find ourselves in. During communication, we use the space around us to establish meaning, particularly through deixis. This can be seen in physical interactions when we use directing-to or placing-for gestures to connect what we are saying to the world around us, but as outlined in the beginning of this thesis, in a changing technological world we need to ask: how do individuals establish deixis in a virtual environment? I believe that this thesis has addressed this overarching question, at least with respect to the specific context of video game tutorials. This chapter will discuss the practical implications of this research, some avenues for the further research, and provide a summary of my findings and discuss their theoretical relevance.

6.1 Practical Implications

As previewed in the introduction, there are several practical implications of this research. The results from this thesis present a typology of the different ways content creators make references and communicate with their audience. This typology may assist or influence game and software developers to create new types of gesturing methods and programs or improve on programs that currently exist. Improvement is essential for software developing companies in that content creators have many types of streaming and recording programs to choose from as they make their videos, OBS, Xsplit, and Bandicam to name a few. As there is competition between the types of programs, content creators choose the one(s) that have the affordances that they are

looking for, such as types of editing features or the ability to present and indicate clearly.

Similarly, software developers have created programs that are ideal for collaborative work, such as Zoom or Microsoft Teams. In collaborative work, deixis is ubiquitous, as people need to be able to draw attention to things on their shared materials and pointing at objects on the screen fulfills this function. In fact, a screen share in Zoom could be considered as a way of achieving joint attention, quite in line with Clark's (2003) placing-for. In addition to using software for collaborative work, in online education, deictic gestures can also play an important part in the making of educational or instructional videos. Knowing how deictic gestures interact with ambiguity and provide clarity can also potentially help those who create the content to be conscious of what and how they display material, and how to best use deictic gestures to make sure they present their information clearly. Therefore, it may be relevant to investigate, aside from the gestures presented in the typology, what other types can be created to help foster communication and collaboration in a working or educational environment. However, with respect to its practical implications, it is important to highlight that my work is entirely observational. While I can show which gestures are used, and in which context, my data does not allow any conclusions about which of these gestures work best. For this, experimental research that assesses people's understanding would be a useful further application of my work.

6.2 Future Research

What I have discussed is only be a starting point. If the gestures presented in my thesis are considered part of the repertoire of human communication, then, as mentioned previously, there is the potential for other new forms of gestures that can be found in other virtual environments. Not only is there potential for further study into other new forms of gestures, but there is potential in expanding this research into other areas of linguistics, such as sociolinguistics, by comparing content created by players from different cultural backgrounds and languages. As has already been established by Kita (2009), in the physical world, different cultures can gesture differently, the question to be answered then is if this extends into the virtual environment: do people from different backgrounds, who are using the same medium, such as YouTube tutorials, communicate differently to their audience? What are the similarities and differences across individuals, cultures, and languages? In my thesis, the content creator T3hasiangod is from the United States, Deligracy is from Australia, and BeastyqtSC2 is from Serbia. Despite their different geographic and cultural space, they shared similar deictic behaviors, such as pointing, circular pointing, and placing-for, along with using mostly proximal deictic expressions. However, for gestures they did not share, such as boxing, rotation, and zoom, were these due solely to the affordance of their game, or are there other, unforeseen, factors? For the present data, differences between individuals are entirely confounded with differences between cultures, as well as differences between games. This limitation is a natural outgrowth of my decision to focus on a detailed qualitative analysis of video material, which involves a very time-consuming annotation process. Future work that takes a different approach may sample

more individuals from different games and cultures and endeavor to disentangle such factors.

The research I laid out in this thesis only examines demonstratives co-occurring with deictic gestures. However, it is still important to mention that nominals also occurred quite frequently with deictic gestures in my data. It would be interesting to investigate this phenomenon in the future. I stated in the literature review that nominals do not necessarily need deictic gestures because the object of reference is stated within the utterance. The question, then, is why there were still several incidents of nominals that co-occurred with deictic gestures in my data.

A further path of exploration would be to emulate the work of Novack et al. (2015) and build research into the reactions and interpretations of the audience. In Novack et al. (2015) they asked participants if the actions of an individual using hand movements were considered communicative, goal-oriented, or movements that were for the sake of movement. Potential future research could be to see if audience members of YouTube gaming tutorials interpret the use of mouse movements in the tutorials similarly. For this, experiments could be developed that take my typology as a starting point and apply similar modulations of Novack et al. (2015) to the parameters of these gestures (e.g., duration, area covered, number of rotations in circular points), to see how they affect the likelihood of a communicative interpretation.

6.3 Summary of Findings and Theoretical Implications

Individuals establish communication in the virtual environment by using the tools available to them, and different virtual environments come with different toolkits. In

video game tutorials, one of the main tools of communication, perhaps the primary one, is the mouse cursor, which features in a variety of different deictic behaviors, depending on local communicative context, which video game is talked about, and on individual differences between content creators. Importantly, however, the mouse cursor also serves a functional purpose that is not directly tied to communication, such as moving things in the game world forward. I used several behavioral indicators to differentiate between communicative and non-communicative uses, such as the presence of demonstratives, or the presence of movements that could not possibly serve a functional purpose (e.g., circular pointing). Other types of gestures, such as boxing, had the dual use of being both functional and communicative.

Having differentiated between communicative and non-communicative uses of the available software tools, I then analyzed the data found in my three YouTube gaming tutorials and formulated a typology of six new gesture types, which include circular indication, pointing, varied indication, boxing, placing-for, and stopping-for. From these six types of gestures, four were used by all three content creators: circular indication, pointing, varied indication, and placing-for. Each individual content creator also had specific referential gestures that were made possible by the features of the programs they used. This helps build expectations for future work on digital deixis. It can be expected that the four most frequent types of deictic virtual behaviors will also occur in other types of games. Moreover, it is reasonable that different software environments will produce different behaviors, and it is possible that for games that are very different from the ones observed here, new types of deictic behaviors could be discovered. The important takeaway from this thesis is that while my typology is

founded on past research, such as Spagnolli, Gamberini and Lucia (2005), it goes beyond their work by not exclusively focusing on pointing with the mouse. My data and the typology that results from it demonstrate that their analysis, while worthwhile, is too limited. Importantly, the typology of gesture types that I developed in this thesis will prove useful for future empirical work on other types of games, content creators, and tutorial formats.

Some specific aspects of my data analysis are worth recapitulating to highlight the diversity of deictic behaviors observed. BeastyqtSC2's game of *Starcraft II* has the feature of boxing with a mouse, which in the research demonstrated qualities of Clark's (2003) 'directing-to', in that it was primarily used to direct the viewer's attention. Deligracy's game was *The Sims 4*, which, more so than *Starcraft II*, allows the user to manipulate the screen through rotation and zooming. I analyzed this as having features of 'placing-for', in that the game world is being manipulated to place something into the viewer's focus of attention. Finally, T3hasiangod's game was *The Elder Scrolls Online* and his unique deictic gesture was stopping-for, which is clearly a direct consequence of this being the only video game tutorial that used pre-recorded footage, thereby allowing the use of the play and pause feature of Windows Media Player. The fact that each content creator had a different type of deictic gesture that was most frequent also attests to the diversity of deictic behaviors. In T3hasiangod's tutorial, circular gesturing occurred the most, in Deligracy's tutorial pointing occurred the most, and in BeastyqtSC2's tutorial the most frequently used gesture was boxing.

My typology was centered on Clark's (2003) theory of placing-for and directing-to, the only framework in gesture studies that explicitly and exclusively focuses on how

attention factors into bodily movements (either by directing attention, or by manipulating the world with respect to attention). When Clark's work was juxtaposed with my analysis of deictic gestures, a number of key similarities were observed between Clark's observations on deixis in the physical world and the YouTube gaming tutorials analyzed here, such as in the temporal profiles of placing-for and directing-to (Table 2.1 in Chapter 2). The work laid out in this thesis is the first to empirically test Clark's dichotomy, and it is furthermore the first to apply it to the virtual environment. The results support his claim that directing-to gestures are shorter than placing-for — even in the virtual environment. However, the research also indicates that there are places where his dichotomy breaks down. For instance, in YouTube gaming tutorials, it is possible for a deictic gesture to simultaneously fall into both categories of directing-to and placing-for. For example, it is possible to use the mouse cursor to direct attention to an object in the game world that is also being manipulated.

I did not focus solely on virtual deictic gestures, but also investigated the linguistic choices that co-occurred with them, with a specific focus on the distinction between "proximal" and "distal" demonstratives. Results show an overwhelming majority of proximal demonstratives, with very few distals. I argued that specifically for games, this observation has three possible explanations. First, ambiguity or precision, as "this" has been found to be more frequent in contexts that afford precise pointing (Cooperrider, 2016), which is the case in gaming tutorials, as the mouse cursor is a device that can point to pixel-level precision. Second, it is possible that distance is an additional factor, given that things in the game world are "close" from the content creator's perspective, as they are looking at a screen that is right in front of them. Third,

many of the objects pointed to in the gaming world were manipulable, which Rocca and Wallentin (2020) found to be a factor to increase proximal use. Interestingly, though only seen in Deligracy's tutorial, it appeared that distals were more often used when an action was completed, and an object was re-referenced by the content creator. Future research needs to investigate whether this observation extends to other games, and whether the general over-representation of proximals is indeed a general feature of communication in this type of context.

Clark's dichotomy was not the only feature that had yet to be tested in the virtual environment. Though there has been exploratory research into the combined use of speech acts and personal and temporal deixis, the relationship between speech acts and spatial deixis co-occurring with deictic gestures has received far less attention, especially in YouTube gaming tutorials. The contribution from this thesis provides examples that show the combined use of spatial deixis with deictic gestures do indeed tie speech acts to their context, and that this area is ripe for further research on the virtual environment.

There were other theoretical similarities between deictic gestures in the virtual environment and those in the physical world. Some examples include Clark's (2003) preparatory principle, Kita's (1993) post-stroke hold, and the use of deictic gestures in the virtual environment to think out loud, retrieve words, and to help with 'internal computation' (Kirsh, 1995). Others have also already discussed mouse cursor movements as gestures (Spagnolli et al., 2005), but my analysis goes deeper, pointing out even more parallels between physical and virtual deictic gestures. This, together with the successful application of Clark's (2003) and Cooperrider's (2016) insights to my

data suggests deep similarities between gestures in the physical world, and gestures in the virtual environment. Indeed, the fact that all of these concepts from gesture research carried over quite nicely to the virtual context makes a strong case for mouse cursor movements being gestures in a genuine sense. The mouse acts as if it is a direct extension of the hand, and the types of manual gestures observed have direct parallels in virtual environments. As most existing work on gestures focuses on manual co-speech gestures (McNeill, 1992; Kendon, 2004), the deep parallels between physical and virtual gestures that form the core of my thesis suggest that virtual gesturing needs to be more directly incorporated into theoretical frameworks in gesture research.

Given how dominant the virtual environment is in our modern everyday lives, we may even say that any theory of gesture is incomplete without considering virtual environments. I am proposing that all the forms of indication presented in my research should not only be considered types of gestures but should also be added to the repertoire of current gesture research. In addition, investigation into how deictic behaviors are used to establish meaning and communication in other parts of the virtual environment is crucial for more deeply understanding everyday communication in this environment.

References

Abdullah, A. (2014). *An Ethnographic Sociolinguistic Study of Virtual Identity in Second Life*. PhD thesis, University of Leeds.

Abdullah, M. (2015). Deixis: A Pragmatics Analysis. *Language in India*, *15*(12), pp. 3-9.

AbdulilahGheni, A (2019). A Study of Deictic Shift Theory as a Stylistic Approach in the Analysis of Point of View Effects in Fictional Discourse. *Al-Ustath Journal for Human and Social Sciences*, *58*(2), pp. 129-142.

Ahlsen, E. (2015). *Gestures Used in Word Search Episodes by Persons with and without Aphasia.* Proceedings of the 2nd European and The 5th Nordic Symposium on Multimodal Communication, Linköping, Sweden.

Aimaiti, N., & Yan, X. (2011). *Gestire-based Interaction and Implication for the Future*, Umeå University.

Alibali, M., & Heath, D. (2001). Effects of Visibility between Speaker and Listener on Gesture Production: Some Gestures Are Meant to Be Seen. *Journal of Memory and Language, 44*, pp. 169-188.

Alibali, M., Young, A., Crooks, N., Yeo, A., Wolfgram, M., Ledesma, I., Nathan, M., Church, R., & Knuth, E. (2013). Students Learn More When Their Teacher Has Learned to Gesture Effectively. *Gesture, 13*(2), pp. 210-233.

Almobarraz, A. (2018). Utilization of YouTube as an Information Resource to Support University Courses. *The Electronic Library, 36*(1), pp. 71-81.

Anderson, M. (2020). The 13 most popular types of videos on YouTube [Infographic]. Impact, viewed 2 Dec 2020. <https://www.impactplus.com/blog/most-popular-types-of-videos-on-youtube-info

graphic>

Andrén, M. (2017). ‘Children’s Expressive Handling of Objects in a Shared World’, in Meyer, C., Streeck, J., and Scott, J. (eds.) *Intercorporeality: Emerging Socialities in Interaction.* Oxford University Press, pp. 105-141.

Andric, M., & Small, S. (2012). Gesture’s Neural Language. *Frontiers in Psychology, 3*, pp. 1-12.

Ang, C. (2020). Who’s the Most Popular YouTuber in Every Country? Visual Capitalist, viewed 2 Dec 2020, <https://www.visualcapitalist.com/worlds-most-popular-youtubers/>

Angouri, J. (2018). ‘Quantitative, Qualitative, Mixed or Holistic Research? Combining Methods in Linguistic Research’, in Litosseliti, L. (ed.) *Research Methods in Linguistics*. London, UK: Bloomsbury, pp. 35-56.

Antonijevic, S. (2008). From Text to Gesture Online: A Microethnographic Analysis of Nonverbal Communication in the Second Life Virtual Environment. *Information, Community and Society, 11*(2), pp. 221-238.

Aslam, S. (2020). YouTube by the Numbers: Stats, Demographics & Fun Facts. Omnicore, viewed 10 Sept 2020, <https://www.omnicoreagency.com/youtube-statistics/>

Austin, J. L. (1962). *How to Do Things with Words*. Cambridge: Harvard University Press.

The British Association for Applied Linguistics (BAAL). (2021). Recommendations on Good Practice in Applied Linguistics. 4th Edition. Available at www.baal.org.uk

Baker, P. (2018). ‘Corpus Methods in Linguistics’, in Litosseliti, L. (ed.) *Research*

Methods in Linguistics. London, UK: Bloomsbury. pp. 167-191.

Bamford, J. (2004). ‘Gestural and Symbolic Uses of the Deictic *here* in Academic Lectures’, in Aijmer, K. and Stenström (eds.) Discourse Patterns in Spoken and Written Corpora: Pragmatics and Beyond New Series 120, John Benjamins Publishing Company, pp. 113-138.

Bartle, R. (2003). *Designing Virtual Worlds*. New Riders Games.

BeastyqtSC2. (2018). StarCraft 2: Learning The Basics - Bronze to Grandmaster - PART 1/7 (Guide by a Pro Player) <https://www.youtube.com/watch?v=FxxgUGBfjRg>

Bell, M. (2008). Toward a Definition of "Virtual Worlds". *Journal of Virtual Worlds Research, 1*(1), pp. 1-5.

Bell, A., Ensslin, A., Van Der Bom, I., & Smith, J. (2018). Immersion in Digital Fiction. *International Journal of Literary Linguistics, 7*(1), pp. 1-22.

Bernard, J., Millman, Z., & Mittal, V. (2015). Beat and Metaphoric Gestures are Differentially Associated with Regional Cerebellar and Cortical Volumes. *Human Brain Mapping, 36*, pp. 4016-4030.

Berry, R. (2009). You Could Say That: The Generic Second-Person Pronoun in Modern English. *English Today, 25*(3), pp. 29-34.

Bhatia, A. (2018). Interdiscursive Performance in Digital Professions: The Case of YouTube Tutorials. *Journal of Pragmatics, 124*, pp. 106-120.

Boxer, D. (2002). Discourse Issues in Cross-Cultural Pragmatics. *Annual Review of Applied Linguistics, 22*, pp. 150-167.

British Psychological Society (2021). Ethics Guidelines for Internet-Meditated Research.

Available at https://www.bps.org.uk/node/1741.

Buchanan, E., & Ess, C. (2008). ‘Internet Research Ethics: The Field and Its Critical Issues’, in Elinar Himma, K., and Tavani, H. (eds), *The Handbook of Information and Computer Ethics*: New Jersey, John Wiley & Sons, Inc., pp. 273-292.

Bühler, K. (1990). *Theory of language : The Representational Function of Language.* John Benjamins.

Burke, S., Snyder, S., & Rager, R. (2009). An Assessment of Faculty Usage of YouTube as a Teaching Resource. *The Internet Journal of Allied Health Sciences and Practice, 7*(1), pp. 1-8.

Butterworth, G. (2003). ‘Pointing Is the Royal Road to Language for Babies’, in Kita, S. (ed), *Pointing: Where Language, Culture, and Cognition Meet*. Psychology Press: Taylor & Francis Group, pp. 9-33.

Cai, S., Zhu, G., Wu, Y. T., Liu, E., & Hu, X. (2018). A Case Study of Gesture-Based Games in Enhancing the Fine Motor Skills and Recognition of Children with Autism. *Interactive Learning Environments, 26*(8), pp. 1039-1052.

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council, Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, (2018). Available at https://ethics.gc.ca/eng/nr-cp_2019-06-05.html.

Carter, M., Gibbs, M., & Arnold, M. (2012). *Avatars, Characters, Players and Users: Multiple Identities at/in Play*. Proceedings of the 24th Australian Computer-Human Interaction Conference (OzCHI '12), Association for Computing Machinery, pp. 68–71.

Chen, M. (2012). *Leet Noobs: The Life and Death of an Expert Player Group in World of Warcraft.* Peter Lang Publishing, New York.

Chen, M., Anderson, J., & Sohn, M. (2001). *What Can a Mouse Cursor Tell Us More? Correlation of Eye/Mouse Movements on Web Browsing*. In CHI '01 Extended Abstracts on Human Factors in Computing Systems (CHI EA '01), Association for Computing Machinery, pp. 281-282.

Chik, A. (2013). Naturalistic CALL and Digital Gaming. *TESOL Quarterly, 47*(4), pp. 834-839.

Choi, G., & Behm-Morawitz, E. (2017). Giving a New Makeover to STEAM: Establishing YouTube Beauty Gurus as Digital Literacy Educators Through Messages and Effects on Viewers. *Computers in Human Behavior, 73*, pp. 80-91.

Ciroux, S. (2020). *How to Say Things with Hand Gestures*. Proceedings of the 24th Workshop on the Semantics and Pragmatics of Dialogue, pp. 1-11

Clark, H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Clark, H. (2003). ‘Pointing and Placing’, in Kita, S. (ed.), *Pointing. Where Language, Culture, and Cognition Meet*. Hillsdale NJ, Erlbaum, pp. 243-268.

Clark, H. (2005). Coordinating With Each Other in a Material World. *Discourse Studies, 7*(4-5), pp. 507-525.

Colletta, J., Guidetti, M., Capirci, O., Cristilli, C., Demir, O., Kunene-Nicolas, R., & Levine, S. (2015). Effects of age and language on co-speech gesture production: An investigation of French, American, and Italian children's narratives. *Journal of Child Language, 42*(1), pp. 122-145.

Connolly, T., Boyle, E., MacArthur, E., Hainey, T., & Boyle, J. (2012). A Systematic

Literature Review of Empirical Evidence on Computer Games and Serious Games. *Computers & Education, 59*, pp. 661-686.

Cooke-Plagwitz, J. (2008). New Directions in CALL: An Objective Introduction to Second Life. *CALICO Journal, 25*(3), pp. 547-557.

Cooperrider, K. (2014). Body-Directed Gestures: Pointing to the Self and Beyond. *Journal of Pragmatics, 71*, pp. 1-16.

Cooperrider, K. (2016). The Co-Organization of Demonstratives and Pointing Gestures. *Discourse Processes, 53*(8). pp. 632-656.

Cooperrider, K., Slotta, J., & Nunez, R. (2018). The Preference for Pointing With the Hand Is Not Universal. *Cognitive Science, 42*, pp. 1375-1390.

Cooperrider, K., Fenlon, J., Keane, J., Brentari, D., & Goldin-Meadow, S. (2021). How Pointing is Integrated into Language: Evidence From Speakers and Signers. *Frontiers in Communication, 6*, pp. 1-13.

Colthup, H. (2018). 'You Were all the World like a Beach to me' The Use of Second Person Address to Create Multiple Storyworlds in Literary Video Games: 'Dear Esther', a Case Study. *International Journal of Transmedia Literacy*, pp. 121-142.

Cornillie, F., Thorne, S., & Desmet, P. (2012). ReCALL special issue: Digital Games for Language Learning: Challenges and Opportunities. *ReCALL, 24*, pp. 243-256.

Coventry, K., Valdes, B., Castillo, A., & Guijarro-Fuentes, P. (2008). Language Within Your Reach: Near-far Perceptual Space and Spatial Demonstratives. *Cognition, 108*, pp. 889-895.

Coventry, K.R., Griffiths, D., & Hamilton, C.J. (2014). Spatial Demonstratives and Perceptual Space: Describing and Remembering Object Location. *Cognitive*

Psychology, 69, pp. 46-70.

Craig, A., Sherman, W., & Will, J. (2009). *Developing Virtual Reality Applications*. Morgan Kaufmann Publisher: Burlington, MA.

Cserzo, D. (2019). *A Nexus Analysis of Domestic Video Chat: Actions, Practices, Affordances, and Mediational Means*. PhD Thesis, Cardiff University.

Čulig, B., Katavić, M., Kuček, J., & Matković, A. (2014). 'The Phenomenology of Video Games: How Gamers Perceive Games and Gaming', in Stobbart, D., and Evans, M. (eds.) *Engaging with Videogames: Play, Theory, and Practice*. UK, Inter-Disciplinary Press, pp. 65-75.

Damude, S., Hoekstra-Weebers, J.E.H.M., van Leeuwen, B.L., & Hoekstra, H.J. (2017). Melanoma Patients' Disease-Specific Knowledge, Information Preference, and Appreciation of Educational YouTube Videos for Self-Inspection, *EJSO, 43*, pp. 1528-1535.

Dehghani, M., Niaki, M., Ramezani, I., & Sali, R. (2016). Evaluating the Influence of YouTube Advertising for Attraction of Young Customers. *Computers in Human Behavior, 59*, pp. 165-172.

Deligracy. (2018). HOW TO PLAY THE SIMS 4: Moving in & Getting Started [Video]. YouTube. <https://www.youtube.com/watch?v=QhC_hcnq0-s&t=52s>

Diessel, H. (2006). Demonstratives, Joint Attention, and the Emergence of Grammar. *Cognitive Linguistics, 17* (4), pp. 463-489.

Duchan, J., Bruder G., & Hewitt, L. (1995). *Deixis in Narrative. A Cognitive Science Perspective*. Hillsdale, NJ: Lawrence Erlbaum.

Due, B., & Toft, T. (2021). Phygital Highlighting: Achieving Joint Visual Attention When

Physically Co-Editing a Digital Text. *Journal of Pragmatics, 177*, pp. 1-17.

Duo, M. (2020). 10 Best Video Hosting Solutions to Consider in 2020 (Free vs. Paid). Kinsta, viewed 16 Nov 2020 <https://kinsta.com/blog/video-hosting/>

Edley, N. & Litosseliti, E. (2018). 'Critical Perspectives on Using Interviews and Focus Groups', in Litosseliti, L. (ed.) *Research Methods in Linguistics*. London, UK: Bloomsbury, pp. 195-225.

Ensslin, A. (2009). 'Respiratory Narrative. Multimodality and Cybernetic Corporeality in 'Physio-cybertext'', in Page, R. (ed.) *New Perspectives on Narrative and Multimodality*. London, Routledge, pp. 155-165.

Ensslin, A. (2012). *The Language of Gaming*. Red Globe Press.

Ensslin, A., & Balteiro, I. (2019). *Approaches to Videogame Discourse: Lexis, Interaction, Textuality*. London: Bloomsbury Academic.

Ensslin, A., & Finnegan, J. (2019). 'Bad Language and Bro-up Cooperation in Co-sit Gaming.' in Ensslin, A. and Balteiro, I. (eds.) *Approaches to Videogame Discourse: Lexis, Interaction, Textuality.* London, Bloomsbury Academic, pp. 139-156.

Eriksson, M. (2009). Referring as Interaction: On the Interplay Between Linguistic and Bodily Practices. *Journal of Pragmatics, 41*, pp. 240-262.

Facca, D., Smith, MJ., Shelley, J., Lizotte, D., Donelle, L. (2020). Exploring the ethical issues in research using digital data collection strategies with minors: A scoping review. *PLoS ONE 15*(8), pp. 1-17.

Fauconnier, G., & Turner, M. (2003). Conceptual Blending, Form and Meaning. *Recherches en Communication, 19: Sémiotique cognitive — Cognitive*

Semiotics, pp. 57-86.

Fauville, G., Luo, M., Queiroz, A. C. M., Bailenson, J. N., & Hancock, J. (2021). *Nonverbal Mechanisms Predict Zoom Fatigue and Explain Why Women Experience Higher Levels than Men.* Available at SSRN: *https://ssrn.com/abstract=3820035*.

Ferchaud, A., Grzeslo, J., Orme, S., & LaGroue, J. (2018). Parasocial Attributes and YouTube Personalities: Exploring Content Trends Across the Most Subscribed YouTube Channels. *Computers in Human Behavior, 80*, p. 88-96.

Feyereisen, P. (2017). *The cognitive psychology of speech-related gesture. Taylor & Francis Group*. London: Routledge.

Fralinger, B., & Owens, R. (2009). YouTube as a Learning Tool. *Journal of College Teaching & Learning, 6*(8), pp. 15-28.

Frank, M., Braginsky, M., Yurovsky, D., & Marchman, V. (2021). *Variability and Consistency in Early Language Learning: The Wordbank Project*. Cambridge, MA: MIT Press.

Fricke, E. (2003). 'Origo, Pointing, and Conceptualization - What Gestures Reveal About the Nature of the Origo in Face-to-Face Interaction', in Lenz, F. (ed.) *Deictic Conceptualisation of Space, Time, and Person*. Pragmatics and Beyond Series, 112. pp. 69-93.

Frobenius, M. (2013). Pointing Gestures in Video Blogs. *Text & Talk, 33*(1), pp. 1-23.

Frobenius, M. (2014). Audience Design in Monologues: How Vloggers Involve Their Viewers. *Journal of Pragmatics, 72*, pp. 59-72.

Gass, S. (1996).). 'Introduction', in Gass, S. and Neu, J. (eds) *Speech Acts Across*

Cultures. Berlin, Mouton de Gruyter, pp. 1-17.

Gee, J. (2007). *What Video Games Have to Teach Us About Learning and Literacy.* Palgrave MacMillan

Gee, J., & Hayes, E. (2012). 'Nurturing Affinity Spaces and Game-Based Learning', in Steinkuehler, C., Squire, K., and Barab, S. (eds.) *Games, Learning, and Society: Learning and Meaning in the Digital Age*. Cambridge: Cambridge University Press. pp. 129-153.

Gelinas, L., Pierce, R., Winkler, S., Cohen, G., Fernandez Lynch, H., & Bierer, B. (2017). Using Social Media as a Research Recruitment Tool: Ethical Issues and Recommendations. *Am J Bioeth, 17*(3), pp. 3-14.

Gerhardt, C. (2019). ''Showing' as a Means of Engaging a Reluctant Participant into a Joint Activity', in Reber, E., and Gerhardt, C. (eds) *Embodied Activities in Face-to-Face and Mediated Settings*. Palgrave Macmillan, pp. 137-175.

Gerring, J. (2017). Qualitative Methods. *Annu. Rev. Polit. Sci. 20*, pp. 15-36.

Ghasemi, B., Hashemi, M., & Bardine, S. (2011). The Capabilities of Computers for Language Learning. *Procedia - Social and Behavioral Sciences, 28*, pp. 58-62.

Gibson, J. (1979). The Ecological Approach to Visual Perception. Boston, MA: Houghton Mifflin.

Ginzburg, J., & Poesio, M,. (2016). Grammar Is a System That Characterizes Talk in Interaction. *Frontiers in Psychology, 7*, pp. 1-22.

Gledhill, C. (2019). 'Phraseology and Lexico-grammatical Patterns in Two Emergent Paragame Genres: Videogame Tutorials and Walkthroughs', in Ensslin, A., and Balteiro, I. (eds.) *Approaches to Videogame Discourse: Lexis, Interaction,*

Textuality. London: Bloomsbury Academic. pp. 58-86.

Glen, S. (2016). "Maximum Variation Sampling" From StatisticsHowTo.com: Elementary Statistics for the rest of us! https://www.statisticshowto.com/maximum-variation-sampling/

Glover, K. (2000). Proximal and Distal Deixis in Negotiation Talk. *Journal of Pragmatics, 32*, pp. 915-926.

Goertzen, M. (2017). Introduction to Quantitative Research and Data. *Library Technology Reports, 53*(4), pp. 12–18.

Goodwin, M., & Goodwin, C. (1986). Gesture and Coparticipation in the Activity of Searching for a Word. *Semiotica, 62*, pp. 51-75.

Goodwin, C. (2003). 'Pointing as Situated Practice', in Kita, S. (ed.) *Pointing. Where Language, Culture, and Cognition Meet*. Hillsdale NJ: Erlbaum. pp. 217-241.

Goodwin, C. (2007). 'Environmentally Coupled Gestures', in Duncan, S., Cassell, J., and Levy, E. (eds) Gesture and the Dynamic Dimensions of Language. Amsterdam: John Benjamins, pp. 195-212.

Goodwin, C. (2013). The Co-Operative, Transformative Organization of Human Action and Knowledge. *Journal of Pragmatics, 46*, pp. 8-23.

Godwin, H., Khan, M., & Yellowlees, P. (2017). The Educational Potential of YouTube. *Acad Psychiatry, 41*, pp. 823-827.

Goel, L., Prokopec, S., & Junglas, I. (2013). Coram Populo - In the Presence of People: The Effect of Others in Virtual Worlds. *Journal of Computer-Mediated Communication*, pp. 265-282.

Goldin-Meadow, S. & Butcher, C. (2003). 'Pointing Toward Two-Word Speech in Young

Children', in Kita, S. (ed.) *Pointing. Where Language, Culture, and Cognition Meet*. Hillsdale NJ: Erlbaum, pp. 85-107.

Golub, A. (2010). Being in the World (of Warcraft): Raiding, Realism, and Knowledge Production in a Massively Multiplayer Online Game. *Anthropological Quarterly, 83*, pp. 17-45.

Granic, I., Lobel, A., & Engels, R. (2014). The Benefits of Playing Video Games. *American Psychological Association, 69*(1), pp. 66-78.

Gullberg, M. (2006). Some Reasons for Studying Gesture and Second Language Acquisition (Hommage à Adam Kendon). *IRAL, 44*, pp. 103-124.

Gupta, S. (2017). Ethical Issues in Designing Internet-Based Research: Recommendations for Good Practice. *Journal of Research Practice, 13*(2), pp. 1-14.

Hamilton, A., & Finley, E. (2020). Reprint of: Qualitative methods in implementation research: An introduction. *Psychiatry Research, 283*, pp. 1-8.

Hanks, W. (2009). Fieldwork on Deixis. *Journal of Pragmatics, 41*, pp. 10-24.

Hanks, W. (2017). *Deixis and Pragmatics*. Oxford Research Encyclopedia of Linguistics.

Hassemer, J. (2009). *Chief Pointing-Eye: Multiarticulatory Metaphorical Gestures Expressing High*. PhD thesis. European University of Frankfurt. Retrieved from https://ssrn.com/abstract=1507135 or http://dx.doi.org/10.2139/ssrn.1507135

Hassemer, J., & McCleary, L. (2018). The multidimensionality of pointing. *Gesture, 17*(3), pp.417-463.

Haverkate, H. (1992). Deictic Categories as Mitigating Devices. *Pragmatics, 2*(4), pp. 505-522.

Haviland, J. (2000). ‘Pointing, Gesture Spaces, and Mental Maps’, in McNeill, D. (ed.) *Language and Gesture*, pp. 13-46.

He, Y., Nagels, A., Schlesewsky, M., & Straube, B. (2018). The Role of Gamma Oscillations During Integration of Metaphoric Gestures and Abstract Speech. *Frontiers in Psychology, 9*, pp. 1-11.

Herman, D. (2002). *Story Logic: Problems and Possibilities of Narrative*. Lincoln, NE: University of Nebraska Press.

Hindmarsh, J., & Heath, C. (2000). Embodied Reference: A Study of Deixis in Workplace Interaction. *Journal of Pragmatics, 32*, pp. 1855-1878.

Hokke, S. Hackworth, N., Quin, N., Bennetts, S., Hnin Y.N., Nicholson, J., Zion, L., Lucke, J., Keyzer, P., Crawford, S. (2018). Ethical issues in using the internet to engage participants in family and child research: A scoping review. *PloS ONE, 13*(9), pp. 1-30.

Holler, J., Kendrick, K. & Levinson, C. (2017). Processing Language in Face-to-Face Conversation: Questions with Gestures Get Faster Responses. *Psychon Bull Review, 25, Springer*, pp. 1900-1908.

Huang, Y. (2014). *Pragmatics*. Oxford Textbooks in Linguistics.

Hull, J., Brunelle, T., Prescott, A., & Sargent, J. (2014). A Longitudinal Study of Risk-Glorifying Video Games and Behavioral Deviance. *Journal of Personality and Social Psychology, 107*(2), pp. 300-325.

Ivankova, N., & Creswell, J. (2009). ‘Mixed Methods’, in Heigham, J., and Croker, R. (eds.) *Qualitative Research in Applied Linguistics: A Practical Introduction*. UK: Palgrave Macmillan, pp. 135-161.

Jang, S. H. (2011). YouTube as an Innovative Resource for Social Science Research. *Proceedings of the Australian Association for Research in Education Conference (AARE 2011 Conference), 1-16.*

Johannes, N., Vuorre, M., & Przybylski, A. (2021). Video Game Play is Positively Correlated with Well-being. *Royal Society Open Science, 8*, pp. 1-14.

Jones, T., & Cuthrell, K. (2011). YouTube: Educational Potentials and Pitfalls. *Computers in the Schools, 28*(1), pp. 75-85.

Jorge, C., van Leeuwen, J.P., Dams, D., & Bouwen, J. (2013). Communicating Deictic Gestures through Handheld Multi-Touch Devices. *Proceedings of the NEM Summit*, pp. 63-68.

Jun, J., Seo, W., Park, J., Park, S., & Jung, H. (2021). Exploring the Experiences of Streamers with Visual Impairments. *Proceedings of the ACM on Human-Computer Interaction, 5*(297), pp. 1-23.

Kapp, K. (2012). *The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education*. Pfeiffer: San Francisco, CA.

Kawano, S., Ono, H., Takagi, T., & Bono, H. (2011). Tutorial Videos of Bioinformatics Resources: Online Distribution Trial in Japan Named TogoTV. *Briefings in Bioinformatics, 13*(2), pp. 258-268.

Kecskes, I. (2013). *Intercultural Pragmatics*. Oxford University Press.

Kellerman, S. (1992). I See What You Mean: The Role of Kinesic Behaviour in Listening and Implications for Foreign and Second Language Learning. *Applied Linguistics, 13*(3), pp. 239-241.

Kelly, S., Healey, M., Ozyurek, A. (2015). The Processing of Speech, Gesture, and

Action During Language Comprehension. *Psychon Bull Rev, 22*, pp. 517-523.

Kendon, A. (1988). ‘Goffman’s Approach to Face-to-Face Interaction’, in Drew, P., and Wooton, A. (eds.) *Erving Goffman: Exploring the interaction order*. La Haye: Mouton, pp. 207-227.

Kendon, A. (1997). Gesture. *Annual Review of Anthropology, 26*, pp. 109-128.

Kendon, A. (2000). ‘Language and gesture: Unity or duality?’ in McNeill, D. (ed.) *Language and Gesture (Language Culture and Cognition)*, pp. 47-63.

Kendon, A., & Versante, L. (2003). ‘Pointing by Hand in “Neapolitan.”’, in Kita, S. (ed.) *Pointing. Where Language, Culture, and Cognition Meet*. Hillsdale NJ: Erlbaum. pp. 109-137.

Kendon, A. (2004). *Gesture: Visible Action as Utterance.* Cambridge: Cambridge University Press.

Kendon, A. (2011). *Gesticulation and Speech: Two Aspects of the Process of Utterance*. The Relationship of Verbal and Nonverbal Communication, Berlin, New York: De Gruyter Mouton, pp. 207-228.

Khan, M. (2017). Social Media Engagement: What Motivates User Participation and Consumption on YouTube? *Computers in Human Behavior, 66*, pp. 236-247.

Kim, J., Kim, M., & Yoo, K. (2013). Real-Time Hand Gesture-Based Interaction with Objects in 3D Virtual Environments. *International Journal of Multimedia and Ubiquitous Engineering, 8*(6), pp. 339-348.

Kirsh, D. (1995). The Intelligent Use of Space. *Artificial Intelligence, 73*, pp. 31-68.

Kistner, J., Dipper, L., & Marshall, J. (2019). The use and function of gestures in word-finding difficulties in aphasia. *Aphasiology, 33*:11, pp. 1372-1392.

Kita, S. (1993). *Language and Thought Interface: A Study of Spontaneous Gestures and Japanese Mimetics*. PhD thesis. University of Chicago.

Kita, S. (2000). 'How Representational Gestures Help Speaking', in McNeill, D. (ed.) *Language and Gesture*, pp. 162-185.

Kita, S. (2003). *Pointing: Where Language, Culture, and Cognition Meet.* Hillsdale NJ: Erlbaum.

Kita, S. (2009). Cross-Cultural Variation of Speech-Accompanying Gesture: A Review. *Language and Cognitive Processes, 24*(2), pp. 145-167.

Kitchin, H. (2003). The Tri-Council Policy Statement and Research in Cyberspace: Research Ethics, The Internet, and Revising a 'Living Document'. *Journal of Academic Ethics, 1*, pp. 397-418.

Klippi, A. (2015). Pointing as an Embodied Practice in Aphasic Interaction. *Aphasiology, 29*(3), pp. 337-354.

Koike, D. (1989). Requests and the Role of Deixis in Politeness. *Journal of Pragmatics, 13*, pp. 187-202.

Krauss, R. (1998). Why Do We Gesture When We Speak? *Current Directions in Psychological Science*, *7*, pp. 54-60.

Krauss, R., Chen, Y., & Gottesman, R. (2000). 'Lexical Gestures and Lexical Access: A Process Model', in McNeill, D. (ed.) *Language and Gesture*, pp. 261-283.

Kurniasih, N., Falaakh, M., & Setyaningsih, N. (2022). Illocutionary Speech Act Analysis of Virtual Youtuber's Viewers in Super Chat. *Undergraduate Conference of Language, Literature, and Culture, 2*(1), pp. 298-308.

Lausberg, H., & Sloetjes, H. (2009). Coding Gestural Behavior with the

NEUROGES–ELAN System. *Behavior Research Methods, 41*(3), pp. 841-849.

Lazzaro, N. (2004). Why We Play Games: Four Keys to More Emotion Without Story. *XEODesign,* pp. 1-8.

Lee, H., Hampel, R., & Kukulska-Hulme, A. (2019). Gesture in Speaking Tasks Beyond the Classroom: An Exploration of the Multimodal Negotiation of Meaning via Skype Videoconferencing on Mobile Devices. *System. 81*, pp. 26-38.

Lee H., Jeong H., Lee J., Yeom KW., & Park, JH. (2009). 'Gesture-Based Interface for Connection and Control of Multi-device in a Tabletop Display Environment', in Jacko J.A. (ed.) *Human-Computer Interaction*. *Novel Interaction Methods and Techniques. HCI 2009. Lecture Notes in Computer Science, vol 5611*. Springer, Berlin, Heidelberg.

Lee, C., Osop, H., Hoe-Lian Goh, D., & Kelni, G. (2017). Making Sense of Comments on YouTube Educational Videos: A Self-directed Learning Perspective. *Online Information Review, 41*(5), pp. 611-625.

Lee, J., & Tatar, D. (2012). "Good Enough" Pointing in Pervasive Computing. *Proceedings of the 2012 International Conference on Collaboration Technologies and Systems, CTS 2012*, pp. 330-337.

Legewie, N., & Nassauer, A. (2018). YouTube, Google, Facebook: 21st Century Online Video Research and Research Ethics. *Qualitative Sozialforschung, 19*(3), pp. 1-22.

Lenz, F. (2003). 'Deictic Conceptualisation of Space, Time and Person', in Lenz, F. (ed.) *Deictic Conceptualisation of Space, Time, and Person.* Pragmatics and Beyond Series, 112, pp. VII-XIV.

Levinson, S. (2006) ‘Deixis’, in Horn, L., and Ward, G. (eds) *The Handbook of Pragmatics*. Blackwell Publishing, UK, pp. 97-121.

Lewandowski, W., & Ozcaliskan, S. (2018). How Event Perspective Influences Speech and Co-Speech Gestures about Motion. *Journal of Pragmatics, 128*, pp. 22-29.

Liang, M. (2021). Pragmatic Socialization Through Gameplay Directives: Multimodal Conversation Analysis of Avatar-Embodied Interactions. *Journal of Pragmatics, 171*, pp. 36-48.

Licoppe, C. (2017). Showing Objects in Skype Video-Mediated Conversations: From Showing Gestures to Showing Sequences. *Journal of Pragmatics, 110*, pp. 63-82.

Lin, Y. L. (2017). Co-occurrence of speech and gestures: A multimodal corpus linguistic approach to intercultural interaction. *Journal of Pragmatics, 117*, pp. 155-167.

Locher, M., Jucker, A., & Berger, M. (2015). Negotiation of Space in Second Life Newbie Interaction. *Discourse, Context, and Media, 9*, pp. 34-45.

Lozano-Blasco, R., Latorre-Martinez, M., & Cortes-Pascual, A. (2021). Analyzing Teens an Analysis from the Perspective of Gamers in YouTube. *Sustainability, 13*, pp. 1-21.

Luciani, A. (2007). ‘Virtual Reality and Virtual Environment’, in Luciani, A., and Cadoz, C. (eds.) *Enaction and Enactive Interfaces: A Handbook of Terms*. ACROE, pp. 299-300.

Lyons, J. (1977). *Semantics.* Cambridge: Cambridge University Press.

MacCallum-Stewart, E. (2014). *Online Games, Social Narratives*. Routledge.

Macey, J., Tyrväinen, V., Pirkkalainen, H., & Hamari, J. (2020). Does esports spectating

influence game consumption? *Behaviour & Information Technology, 41*, pp. 181-197.

MacLeod, T. (2009). *Gesture Signs in Social Interaction: How Group Size Influences Gesture Communication*. PhD thesis, University of Glasgow. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.670.6901&rep=rep1&type=pdf

Malliet, S. (2007). Adapting the Principles of Ludology to the Method of Video Game Content Analysis. *The International Journal of Computer Game Research, 7*(1), pp. 1-17.

Mallon, B., & Lynch, R. (2014). Stimulating Psychological Attachments in Narrative Games: Engaging Players with Game Characters. *Simulation & Gaming, 45*(4-5), pp. 508-527.

Markham, A. (2012). Fabrication as Ethical Practice. *Information, Communication & Society, 15*(3), pp. 334-353.

Masataka, N. (2003). 'From Index-finger Extension to Index-finger Pointing: Ontogenesis of Pointing in Preverbal Infants', in Kita, S. (ed.) *Pointing: Where Language, Culture, and Cognition Meet.* Psychology Press: Taylor & Francis Group.

Matsumoto, Y. & Canagarajah, S. (2020). The Use of Gesture, Gesture Hold, and Gaze in Trouble-in-Talk Among Multilingual Interlocutors in an English as a Lingua Franca Context. *Journal of Pragmatics, 169*, pp. 245-267.

McKay, S. (2009). 'Introspective Techniques', in Heigham, J., and Croker, R. (eds)

Qualitative Research in Applied Linguistics: A Practical Introduction. UK: Macmillan, pp. 220-241.

Merriam-Webster. (n.d.). Virtual reality. In Merriam-Webster.com dictionary. Retrieved October 24, 2020, from https://www.merriam-webster.com/dictionary/virtual%20reality

Mesh, K., Cruz, E., van de Weijer, J., Burenhult, N., & Gullberg, M. (2021). Effects of Scale on Multimodal Deixis: Evidence From Quiahije Chatino. *Frontiers in Psychology, 11*, pp. 1-20.

Mey, J. (2007). *Pragmatics: An Introduction*. Blackwell Publishing.

McGonigal, J. (2012). *Reality Is Broken: Why Games Make Us Better and How They Can Change the World.* Penguin Books.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press.

McNeill, D. (2000). *Language and Gesture.* Language Culture and Cognition Series. Cambridge: Cambridge University Press.

McNeill, D. (2006). Gesture: A Psycholinguistic Approach. *The Encyclopedia of Language and Linguistics*, pp. 58–66. https://mcneilllab.uchicago.edu/pdfs/gesture.a_psycholinguistic_approach.cambridge.encyclop.pdf

Miyake, H., & Sugimura, S. (2017). The Effect of Directive Words on Integrated Comprehension of Speech and Iconic Gestures for Actions in Young Children. *Infant and Child Development, 27*(5), pp. 1-9.

MMO Populations, n.d, *The Elder Scrolls Online*, MMO Populations, viewed on 09 Dec

2020 <https://mmo-population.com/r/elderscrollsonline>

Mondada, L. (2006). ‘Video Recording as the Reflexive Preservation and Configuration of Phenomenal Features for Analysis’, in Knoblauch, H., Schnettler, J. R., and Soeffner, H. G. (eds.) *Video-Analysis: Methodology and Methods*. Bern: Lang, pp. 51-68.

Morford, J. & Goldin-Meadow, S. (1997). From Here and Now to There and Then: The Development of Displaced Reference in Homesign and English. *Child Development, 68*(3), pp. 420-435.

Nash, C. (2007). ‘Gestural Regulators in French, Japanese and American English Dialogue’, in Grein M., and Weigand, E. (eds.) *Dialogue and Culture*. Amsterdam: John Benjamins.

Newon, L. (2011). ‘Multimodal Creativity and Identities of Expertise in the Digital Ecology of a World of Warcraft Guild’, in Thurlow, C., and Mroczek, K. (eds.) *Digital Discourse: Language in the New Media*. Oxford Scholarship Online, pp. 1-28.

Nickel, K. & Stiefelhagen, R. (2003). Pointing Gesture Recognition based on 3D-Tracking of Face, Hands and Head Orientation. *Proceedings of the 5th International Conference on Multimodal Interfaces* (*ICMI’03)*, pp. 140-146.

Nordquist, R. (2019). *Definition and Examples of Conceptual Blending*. ThoughtCo. thoughtco.com/what-is-conceptual-blending-cb-1689780.

Nordvall, M. (2012). *Communication in Games*. Bachelor’s thesis. Linköping University, Linköping Sweden.

Novack, M., Wakefield, E., & Goldin-Meadow, S. (2016). What makes a movement a

gesture? *Cognition, 146*, pp. 339-348.

Novack, M., & Goldin-Meadow, S. (2017). Gesture as Representational Action: A Paper about Function. *Psychon Bull Rev 24*, pp. 652–665.

Oleksik, G., Milic-Frayling, N., & Jones, R. (2014). Touch and Gesture: Mediating Content Display, Inscriptions, and gestures across multiple devices. *Pers Ubiquit Comput*, *18*, pp. 1243-1257.

Orus, C., Barles, M., Belanche, D., Casalo, L., Fraj, E., & Gurrea, R. (2016). The Effects of Learner-generated Videos for YouTube on Learning Outcomes and Satisfaction. *Computers & Education, 95*, pp. 254-269.

Palinkas, L., Horwitz, S., Green, C., Wisdom, J., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Adm Policy Ment Health, 42*, pp. 533-544.

Page, R., Barton, D., Unger, J., & Zappavigna, M. (2014). *Researching Language and Social Media: A Student Guide*. Routledge.

Parvizi, J., Chakravarty, R., Og, B., & Rodriguez-Paez, A. (2008). Informed consent: is it always necessary? *Injury, 39*(6), pp. 651-655.

Patton, M. (2015). *Qualitative Research & Evaluation Methods*. Sage Publications.

Pecay, R. (2017). YouTube Integration in Science Classes: Understanding Its Roots, Ways and Selection Criteria. *The Qualitative Report, 22*(4), pp. 1015-1030.

Perveen, A. (2016). Synchronous and Asynchronous E-Language Learning: A Case Study of Virtual university of Pakistan. *Open Praxis, 8*(1), pp. 21-39.

Peterson, M. (2010). Computerized Games and Simulations in Computer-Assisted

Language Learning: A Meta-Analysis of Research. *Simulation & Gaming, 41*(1), pp. 72-93.

Petrova, E. & Gross, N. (2017). 4 reasons people watch gaming content on YouTube. Think with Google, viewed 1 Dec 2020. <https://www.thinkwithgoogle.com/marketing-strategies/video/statistics-youtube-gaming-content/>

Piirainen-Marsh, A., & Tainio, L. (2014). Asymmetries of Knowledge and Epistemic Change in Social Gaming Interaction. *The Modern Language Journal, 98*(4), pp. 1022-1038.

Piwek, P., Jan Beun, R., & Cremers, A. (2007). 'Proximal' and 'distal' in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics, 40*, pp. 694-718.

Plesner, U., & Phillips, L. (2014). *Researching Virtual Worlds: Methodologies for Studying Emergent Practices.* Routledge: New York, NY.

Poche, E., Jha, N., Williams, G., Staten, J., Vesper, M., & Mahmoud, A. (2017). Analyzing User Comments on YouTube Coding Tutorial Videos. *Proceedings of the 2017 IEEE 25th International Conference on Program Comprehension (ICPC)*. pp. 196-206.

Potts, A. (2015). 'LOVE YOU GUYS (NO HOMO)'. *Critical Discourse Studies, 12*(2), pp. 163-186.

Rasinger, S. (2018). 'Quantitative Methods: Concepts, Frameworks and Issues', in Litosseliti, L. (ed.) *Research methods in linguistics*. London, UK: Bloomsbury, pp. 117-138.

Rocca, R., & Wallentin, M. (2020). Demonstrative Reference and Semantic Space: A Large-Scale Demonstrative Choice Task Study. *Frontiers in Psychology, 11*, pp. 1-10.

Rosborough, A. (2016). Understanding Relations Between Gesture and Chronotope: Embodiment and Meaning-Making in a Second-Language Classroom. *Mind, Culture, and Activity, 23*(2), pp. 124-140.

Rosenberg, A. (2010). Virtual World Research Ethics and the Private/Public Distinction. *International Journal of Internet Research Ethics, 3,* pp. 23-37.

Ryu, D. (2013). Play to Learn, Learn to Play: Language Learning through Gaming Culture. *ReCall, 25*(2), pp. 286-301.

Schroeder, R. (1996). *Possible Worlds*: *The Social Dynamic of Virtual Reality Technologies.* Boulder: Westview Press.

Schroeder, R. (2008). Defining Virtual Worlds and Virtual Environments. *Journal of Virtual Worlds Research, 1*(1), pp. 2-3.

Searle, J. (1976). A Classification of Illocutionary Acts. *Language in Society, 5,* pp. 1-23.

Seay, A., Jerome, W., Lee, K., & Kraut, R. (2004). Project Massive: A Study of Online Gaming Communities. *In Extended Abstracts of the 2004 Conference on Human Factors in Computing Systems, CHI 2004,* pp. 1421-1424.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), pp. 701–703.

Shoufan, A. (2018). Estimating the cognitive value of YouTube's educational videos: A learning analytics approach. *Computers in Human Behavior, 92,* pp. 450-458.

Singer, M., Radinsky, J. & Goldman, S. (2008). The Role of Gesture in Meaning Construction. *Discourse Processes, 45*(4-5), pp. 365-386.

Sixsmith, J., & Murray, C. (2001). Ethical Issues in the Documentary Data Analysis of Internet Posts and Archives. *Qualitative Health Research, 11*(3), pp. 423-432.

Sjöblom, M., Törhönen, M., Hamari J., & Macey, J. (2019). The Ingredients of Twitch Streaming: Affordances of Game Streams. *Computers in Human Behavior, 92*, pp. 20-28.

Spagnolli, A., Gamberini, L., & Reno, L. (2005). Mediated Pointing and the Corporeal Field. Intellectica. Revue de l'Association Pour la Recherche Cognitive. *Espace, Interaction & Cognition, 41-42*, pp. 55-73.

Spagnolli, A., Lombard, M., & Gamberini, L. (2009). Mediated Presence: Virtual Reality, Mixed Environments and Social Networks. *Virtual Reality, 13*, pp. 137–139.

Subramani, R. (2010). Insight through Body Language and Non-verbal Communication References In Tirukkural. *Language in India, 10,* pp. 208-218.

Sugiura, L., Wiles, R., & Pope, C. (2017). Ethical Challenges in Online Research: Public/Private Perceptions. *Research Ethics, 13*(3-4), pp. 184-199.

Stam, G., & McCafferty, S. (2008). 'Gesture Studies and Second Language Acquisition: a Review', in McCafferty, S., and Stam, G. (eds.) *Gesture: Second Language Acquisition and Classroom Research*. Routledge: New York, pp. 3-24.

Stapleton, A. (2017). Deixis in Modern Linguistics. *Essex Student Journal, 9*(1), pp. 1-10.

Statista. (2021). Average YouTube Video Length as of December 2018, by category (in minutes). https://www.statista.com/statistics/1026923/youtube-video-category-

average-length/

Straits Research. (2020). Top 10 Most Popular Gaming Genres in 2020. https://straitsresearch.com/blog/top-10-most-popular-gaming-genres-in-2020/

Stevanovic, M., & Monzoni, C. (2016). On the Hierarchy of Interactional Resources: Embodied and Verbal Behavior in the Management of Joint Activities with Material Objects. *Journal of Pragmatics, 103,* pp. 15-32.

Stukenbrock, A. (2014). Pointing to an 'empty' space: Deixis am Phantasma in face-to-face interaction. *Journal of Pragmatics, 74,* pp. 70-93.

Sveningsson Elm, M. (2008). 'How Do Various Notions of Privacy Influence Decisions in Qualitative Internet Research?', in Markham, A., and Baym, N. (eds) *Internet Inquiry: Dialogue Among Researchers*, Thousand Oaks: SAGE. pp. 69-98.

Tashakkori, A., & Creswell, J. (2007). The New Era of Mixed Methods. *Journal of Mixed Methods Research, 1*(3), pp. 3-7.

T3hasiangod. (2018). Trials Boss Guide - Cloudrest - Siroria [Video]. YouTube. <https://www.youtube.com/watch?v=VbDIB1e6tPU>

Thorne, S.L., & Black, R. W. (2007). New Media Literacies, Online Gaming, and Language Education. *CALPER Working Paper Series, No. 8*. The Pennsylvania State University: Center for Advanced Language Proficiency Education and Research.

Tolson, A. (2010). A New Authenticity? Communicative Practices on YouTube. *Critical Discourse Studies, 7*(4), pp. 277-289.

Trommer, A. K. (2011). Wondering About the Intersection of Speech Acts, Politeness

and Deixis: I Wondered and I Was Wondering in the BNC. *ICAME Journal, 35*, pp. 185-204.

University of Birmingham, Code of Practice, September 2021. <https://www.birmingham.ac.uk/documents/university/legal/research.pdf>

Wadley, G., Carter, M., & Gibbs, M. (2015). Voice in Virtual Worlds: The Design, Use, and Influence of Voice Chat in Online Play. *Human-Computer Interaction, 30*(3-4), pp. 336-365.

Walther, J. B. (2002). Research Ethics in Internet-Enabled Research: Human Subjects Issues and Methodological Myopia. *Ethics and Information Technology, 4*, pp. 205-216.

Walther, J.B., Loh, T. & Granka, L. (2005) Let me count the ways. The Interchange of Verbal and Nonverbal Cues in Computer Mediated and Face-to-Face Affinity. *Journal of Language and Social Psychology, 24*(1), pp. 36-65.

Wende, M., Bulut, S., Giese, T., & Anderl, R. (2020). Online Lectures: Strategies for Applying Advanced Digital Media for Future Higher Education. *Proceedings of the 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, pp. 311-318

Wesn (2013). *Deixis.png,* Wikipedia, viewed 10 Sept 2020 <https://commons.wikimedia.org/wiki/File:Deixis.png>

Wessel-Tolvig, B., & Paggio, P. (2016). Revisiting the Thinking-for-Speaking hypothesis: Speech and Gesture Representation of Motion in Danish and Italian. *Journal of Pragmatics, 99,* pp. 39-61.

Whiteman, N. (2012). *Undoing Ethic: Rethinking Practice in Online Research*. Springer.

Whitton, M. (2003). Making Virtual Environments Compelling. *Communications of the ACM, 46*(7), pp. 40-46.

Widlok, T. (2015). 'Ethnicity as Social Deixis', in University of Cologne Forum »Ethnicity as a Political Resource« (ed.). *Ethnicity as a Political Resource: Conceptualizations across Disciplines, Regions, and Periods,* pp. 85-96.

Williamson, B. (2009). Computer Games, Schools, and Young People: A Report for Educators on Using Games for Learning. *Futurelab*, pp. 1-42.

Winter, B. (2019). *Statistics for linguists: An Introduction Using R*. New York: Routledge.

Winter, B., Perlman, M., & Matlock, T. (2013). Using Space to Talk and Gesture about Numbers. *Gesture 13*(3), pp. 377-408.

Won, D. (2019). *Digital Games as Interactive Narrative and Their Eversion: The Advent of Future Narrative*. Master's thesis. Texas Woman's University, Denton, Texas.

Wu, Y. (2004). *Spatial Demonstratives in English and Chinese: Text and Cognition*. John Benjamins Publishing.

Yee, N. (2006). Motivations for Play in Online Games. *Cyberpsychology & Behavior, 9*(6), pp. 772-775.

YouTube (n.d.). *YouTube for Press*, viewed 16 Nov 2020, <https://www.youtube.com/intl/en-GB/yt/about/press/>

YouTube Terms of Service (n.d.). *Terms of Service*, viewed 17 Mar 2021, < https://www.youtube.com/static?template=terms>

Yule, G. (1996). *Pragmatics*. Oxford University Press.

Yulfi, Y. (2017). An Analysis of Deixis and Speech Act used in English Teaching and

Learning Process. *Linguistic, English Education and Art (LEEA) Journal, 1*(1), pp. 82-90.

Zappavigna, M. (2019). The Organized Self and Lifestyle Minimalism: Multimodal Deixis and Point of View in Decluttering Vlogs on YouTube. *Multimodal Communication, 8*(1), pp. 1-14.

Zhao, X., Li, X., Pang, C., Sheng, Q., Wang, S., & Ye, M. (2014). Structured Streaming Skeleton – A New Feature for Online Human Gesture Recognition. *ACM Trans. Multimedia Comput. Commun. Appl., 11*(22), pp. 1-18.

Appendix A

t3hasiangod Today at 12:44 PM

Yeah, you can use screenshots and video for data.

Also RE: image of face, that's up to what your IRB says is ok. I don't mind either way

Appendix B

From: enquiries@deligracy.com <enquiries@deligracy.com>
Sent: Sunday, August 30, 2020 10:51:28 AM
To: Rachel Manley (PhD Eng Lang+App Lin 6 DL Mod)
Subject: Re: Permission to use your content in my dissertation

Hi there Rachel,

Thanks for thinking of my videos as research for your thesis, we would be thrilled for you to use example screenshots from my content.

Best of luck for your thesis and PHD—I would love to read some of it if you are up for sharing!

Have a lovely weekend,
Deli and team

—
The Deligracy Team

Appendix C

[redacted]@gmail.com

Tue 9/1/2020 5:10 PM

To: Rachel Manley (PhD Eng Lang+App Lin 6 DL Mod) <RXM566@student.bham.ac.uk>;

Go right ahead Rachel! You can use it as you deem necessary :D Thank you for asking