UNIVERSITY OF CALIFORNIA,
IRVINE

Generic Statements as a Pedagogical Tool

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Cognitive Sciences

by

Jeff Coon

Dissertation Committee:
Professor Barbara W. Sarnecka, Chair
Professor Michael D. Lee
Assistant Professor Gregory Scontras

2023

DEDICATION

For Michelle. I'm excited for the next adventure. Thanks for making this one worthwhile.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I could only have made it to this point with considerable help. I can never adequately thank everyone, but I will do my best in the space available:

Thank you to my committee members, who have each had a huge impact on my professional and personal development. Barbara Sarnecka gave me remarkable freedom in choosing the sort of research I wanted to pursue. In the process, she helped me develop a workflow that has given me a degree of organization and productivity that would baffle anyone who knew me before graduate school. She has also mentored me in mentoring, such that I am now well equipped to pursue my primary goal in attending graduate school in the first place, which was to teach aspiring researchers. As my co-advisor, Michael Lee has helped turn my enthusiasm for Bayesian methods into something approaching competence. Furthermore, he has been invaluable in helping me find ways to be both happy and productive. I left every meeting with Michael feeling better about myself and my situation than when I went in. Greg Scontras also deserves to be thought of as my co-advisor, considering all the support he has given me. When I first became interested in studying generic statements, he was the one who helped me get acquainted with the literature. Given how I am not even in his department, let alone in his lab, his generosity with his time, advice, and ideas has been astonishing.

Thank you to my co-authors, Alex Etz and Paulina Silva, and Holly Westfall. Alex has been an incredible mentor, colleague, and friend, although he does debase himself by using overpowered DLC characters in Super Smash. Paulina and Holly came to UCI at the same time as I did, and they have always been there to help me think through ideas, celebrate the highs, and recover from the lows.

Thank you to Emily Sumner and Nick Ichien. Emily mentored me through my formative years of graduate school, helping me find my feet as a graduate student. Nick has been a wonderful friend and conversation partner ever since we met during recruitment visits, and his contagious enthusiasm for fun ideas has helped sustain me through the slog.

Thank you to all of the fantastic colleagues in and around my cohort, the Sarnecka Lab, and the Lee Lab. In particular, thank you to Ashley Thomas, Nicole Scalise, Jessica Gonzalez, Nadia Chernyak, Nora Harhen, Bobby Thomas, Nidhi Banavar, Lauren Montgomery, Manuel Villareal, Adriana Chavez de la Pena, Helen Braithwaite, Joachim Vandekerckhove, Mark Steyvers, Jenny Sun, and Priyam Das. Each of you deserves your own paragraph. Getting to know such kind and brilliant people has been one of the most enjoyable parts of the graduate school experience. Also, thank you to Clara Schultheiss and MyLee Ryan-Glass for patiently helping me navigate the bureaucratic aspects of graduate school.

Thank you to the talented undergraduates who assisted in this research: Jalen Jones, Jonah Pryzbezewski, Desiree Sabri, Natalie Beadle, Adeh Hakobian, Albert Le, and Vivian Nguyen.

Above all, thank you to my wife Michelle for being relentlessly supportive and incredibly patient throughout the entire process.

Finally, thank you to the UC open access policy for allowing me to include Chapter 1 in this dissertation after having already published it in *Games & Culture*.

VITA

Jeff Coon

EDUCATION

Doctor of Philosophy in Cognitive Sciences **2023**
University of California, Irvine *Irvine, CA*

Master of Education in Curriculum and Teaching **2015**
University of Oregon *Eugene, OR*

Bachelor of Arts in Biology **2011**
Colorado College *Colorado Springs, CO*

TEACHING EXPERIENCE

Teaching Associate **2021–2022**
University of California, Irvine *Irvine, CA*

Biology Teacher **2015–2018**
Mt. Tahoma High School *Tacoma, WA*

Student Teacher **2014–2015**
Sheldon High School & Kennedy Middle School *Eugene, OR*

Instructor **2014**
Hancock Field Station, Oregon Museum of Science & Industry *Fossil, OR*

Instructor **2011-2013**
Pali Institute *Running Springs, CA*

REFEREED JOURNAL PUBLICATIONS

Esports as a cultural microcosm for studying psycholinguistics **2023**
Games & Culture

Doctoral writing workshops: A preregistered, randomized controlled trial. **2022**
Innovative Higher Education

A Bayesian method for measuring risk propensity in the Balloon Analogue Risk Task **2021**
Behavior Research Methods

Model-based Wisdom of the Crowd for sequential decisions **2021**
Cognitive Science

REFEREED CONFERENCE PUBLICATION

Adjusting the Use of Generalizations Based on Audience Expertise **July 2022**
Proceedings of the 44th Annual Conference of the Cognitive Science Society

Experts interpret generalizations differently than novices **July 2021**
Proceedings of the 43rd Annual Conference of the Cognitive Science Society

HONORS & AWARDS

Pedagogical Fellow, Division of Teaching Excellence and Innovation **2022**
University of California, Irvine

Spring Graduate Fellow, Division of Teaching Excellence and Innovation **2022**
University of California, Irvine

Certificate of Teaching Excellence **2022**
University of California, Irvine

John I. Yellott Scholar Award **2020**
University of California, Irvine

School of Social Sciences Merit Fellowship **2018**
University of California, Irvine

Graduate Dean Recruitment Fellowship **2018**
University of California, Irvine

ABSTRACT OF THE DISSERTATION

Generic Statements as a Pedagogical Tool

By

Jeff Coon

Doctor of Philosophy in Cognitive Sciences

University of California, Irvine, 2023

Professor Barbara W. Sarnecka, Chair

Generic statements are a fundamental linguistic tool for efficiently passing along information. Yet they can leave a lot unsaid, such as how broadly they apply. In this dissertation, I examine how generic statements work as a pedagogical tool. To do so, I use the esport *League of Legends* as a naturalistic experimental setting in which to observe how people with varying degrees of expertise can use generic statements to teach and learn from each other.

In Chapter 1, I discuss the decision to use an esport as an experimental setting. Esports offer many benefits for studying psycholinguistics because they provide complex yet relatively controlled environments, and many people are highly motivated to learn and talk about such environments. Yet we must be somewhat cautious about extending findings to a general population because players are inherently self-selecting.

In Chapter 2, I describe a study investigating whether novices and experts interpret generic statements differently. I find that experts interpret generic statements flexibly, incorporating their prior knowledge as appropriate. In contrast, novices interpret generic statements unsystematically, varying their responses without any discernible pattern. Novices also tend to interpret generic statements as applying more broadly than experts do.

In Chapter 3, I describe a study comparing how listeners interpret generic statements to

how speakers use generic statements. I find that novice listeners tend to overestimate expert speakers' beliefs about how broadly a generic statement applies, while expert listeners do not.

In Chapter 4, I describe a study investigating whether expert speakers might account for the differences between novice and expert listeners by adjusting their use of generic statements based on the expertise of their audience. In the process, I discuss a descriptive model that quantifies expert speakers' behavior in terms of a threshold at which they become willing to make a generic statement. I find that expert speakers do not meaningfully alter that threshold based on their audience.

In Chapter 5, I conclude by highlighting the themes of this dissertation and by discussing opportunities for future research.

INTRODUCTION

How would you explain mosquitoes to someone who had never encountered them before? You might use generic statements, like “mosquitoes are insects,” “mosquitoes suck people’s blood,” and “mosquitoes carry malaria.” Such statements are efficient because they reference the category of mosquitoes in general (Carlson, 1977), allowing the listener to make inferences about previously unencountered mosquitoes (Gelman, Star, & Flukes, 2002).

Such generic statements are powerful pedagogical tools because they can distill complex considerations down into a simple statements of fact. As such, they are particularly prevalent in child-directed speech (Gelman et al., 1998; Pappas & Gelman, 1998; Gelman, Goetz, Sarnecka, & Flukes, 2008). Children seem to be aware of their pedagogical value because they actively seek out such generic statements (Cimpian & Park, 2014) and privilege them in memory (Cimpian & Erickson, 2012). Indeed, generic statements are so fundamental to teaching that they have been referred to as “natural pedagogy” (Csibra & Gergely, 2009).

Such generic statements are also remarkably flexible: “mosquitoes are insects” applies to every mosquito, but only female mosquitoes suck people’s blood, and very few carry malaria. A challenge for any listener interpreting such statements is to figure out how broadly each applies. This flexibility creates the potential for miscommunication, as a listener might interpret a generic statement as applying to a greater or lesser extent than what a speaker had in mind.

It appears that listeners overcome this challenge by using their prior knowledge to restrict potential interpretations. For example, a listener who has prior knowledge about mosquitoes and malaria is unlikely to think that "mosquitoes carry malaria" applies to the entire category. As long as the speaker and listener have comparable prior knowledge, miscommunication should be minimal.

Yet when generic statements are used to teach people, a speaker who knows a fact is trying to pass on their knowledge to a listener who does not. The listener needs prior knowledge to interpret how broadly the generic statement applies, but the generic statement is introducing them to a fact that they did not previously know. Even so, we can safely assume that generic statements are somehow quite effective in teaching people, given their ubiquity. In this dissertation, I work towards understanding how generic statements work as pedagogical tools. I investigate how listeners use their prior knowledge to interpret generic statements, whether there is miscommunication when an expert speaker uses a generic statement to teach a novice listener, and how expert speakers decide when to use generic statements given the expertise of their audience.

0.1 Operational Definition of Miscommunication

In examining the possibility that there might be miscommunication between expert speakers and novice listeners, I am adopting a specific and somewhat unusual operational definition of the term. It is thus worth clarifying a few points about this operational definition.

First, I am focused on whether one specific aspect of a generic statement's message, namely how broadly it applies, is communicated effectively. When a generic statement fails to effectively communicate the prevalence of the referenced trait, I consider this miscommunication. But I do not mean to imply that the statement has necessarily been counterproductive from a pedagogical perspective. For example, if a listener was unaware that mosquitoes could carry malaria, a generic statement that makes them aware of the link between mosquitoes and malaria may well improve their overall understanding of the world, regardless of how broadly they think the statement applies. Yet if a listener misinterprets the implied prevalence of a generic statement, that mistake can have important consequences. In the context of the statement that "mosquitoes carry malaria," misinterpreting the implied prevalence could lead listeners to be excessively or insufficiently cautious.

Second, when I am considering whether a generic statement used in a pedagogical context is communicating effectively or instead causing miscommunication, I am not expecting perfection. In principle, I could construe any situation in which a listener cannot perfectly recreate the speaker's conception of a generic statement's implied prevalence as an example of miscommunication. However, such a standard would be unreasonable. Generic statements are efficient pedagogical tools because speakers can compress information about the prevalence of a trait into a simple statement, rather than explicitly specifying their precise understanding of the prevalence. Listeners then need to decompress the generic statement to infer the implied prevalence. Such decompression is presumably a noisy process, so I would not expect to see expert listeners perfectly recreate an expert speaker's understanding, let alone a novice listener. Since prior knowledge evidently plays an important role in the decompression, I am investigating whether differences in prior knowledge between experts and novices will lead to consistent distortions in what novice listeners infer from a generic statement relative to what expert speakers believe and what expert listeners infer.

The third point to clarify is that I do not mean to imply that such miscommunication should—or even can—be avoided. Instead, I am looking for patterns that have interesting implications for understanding the process by which humans learn. How best to respond to such patterns in terms of specific interventions is beyond the scope of the research presented in this dissertation.

0.2 Experimental Setting

In studying these considerations, I use the esport *League of Legends* as a naturalistic setting in which to observe how people with varying degrees of expertise use and interpret generic statements. This choice is somewhat unusual, so I use Chapter 1 to discuss the possibilities and implications of using an esport as an experimental setting.

Since the focus of this dissertation is on how people with different levels of expertise use and interpret generic statements, I will refer to expert and novice participants throughout the dissertation. In doing so, I am defining experts as such relative to novices; experts are already familiar with the topic under discussion while novices are not. I do not mean to suggest that expert participants are supremely skilled or knowledgeable players, although some of them evidently were. I also do not mean to suggest that novice participants were learning about the game, or even that they were particularly interested in learning about the game, although some of them may well have been.

0.3 The Semantics of Generic Statements

The specific statements I focus on in this dissertation are technically habitual statements, which are generalizations about a single entity over time (e.g., "Michael bikes to work"). In contrast, the generalizations studied in the above-cited research are more conventional generic statements, which are generalizations over category members (e.g., "dogs have fur"). Habitual statements are considered a subtype of generic statement (e.g., Carlson, 2006). Since this dissertation builds from and pertains to work that focused on generic statements, I refer to both types of generalizations simply as "generic statements." I choose to focus on habitual statements because they are more convenient in the context of using an esport as an experimental setting.

In terms of challenge they present for communication, the semantics for habituals and classic generic statements are virtually identical; the statement implies some degree of prevalence, but how much is unclear (Tessler & Goodman, 2019a). For example, "Michael bikes to work" clearly indicates that the referenced person bikes to work in at least some instances, but how often do they need to bike to work for the habitual to be true?

There are currently two leading accounts, both of which necessitate some pragmatic inference on the part of the listener interpreting a generic statement. The first account treats generic statements as representing "default" generic statements and thus focuses on identifying traits that we would naturally generalize to an entire category (Leslie, 2012). More specifically, this account proposes that a generic statement is considered true (i.e., would be endorsed by a reasonable speaker) if the relationship between referenced trait and category meet any of a list of criteria (Leslie, 2008). For example, the trait could be particularly characteristic of the category ("Cardinals are red"), particularly dangerous and thus important to the listener ("mosquitoes carry malaria"), or true more often than not ("cars have radios"). If these criteria determine when a speaker would endorse a statement, then interpretation would presumably involve listeners trying to pragmatically figure out which criteria prompted the speaker to make the statement. This account thus implies that a listener would need to use their prior knowledge to narrow down which criteria might be true. For example, if a listener knows that malaria is a dangerous disease, they might infer that a speaker chose to say that "mosquitoes carry malaria" because it meets the criteria of referencing a particularly dangerous trait.

What I call the second account is more a family of accounts, all focused on the proposition that the implied prevalence of a generic statement is central to its meaning (Cohen, 1999; Tessler & Goodman, 2019a; Van Rooij & Schulz, 2020). The most explicit model of this account (Tessler & Goodman, 2019a) proposes that reasonable speakers will endorse a generic statement if it would improve the expectations of a listener (i.e., make them align more

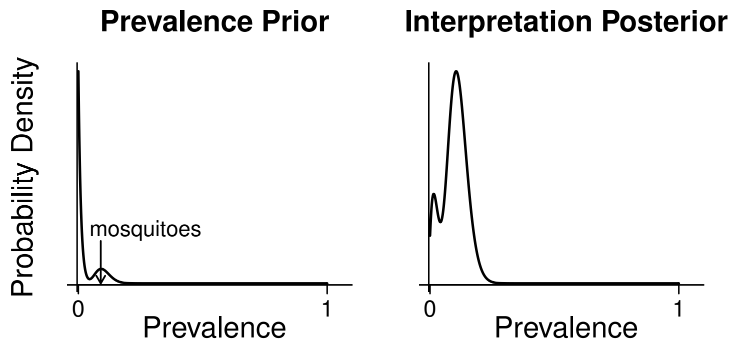

Figure 1: Illustrative graphs inspired by a figure in Tessler and Goodman, 2019a (p. 401). **Left:** Prior expectations of a hypothetical novice regarding the potential prevalence of the trait "carries malaria." Note that the hypothetical novice thinks it most likely that no category members carry malaria, but allows for the possibility that a small minority do. The arrow labeled "mosquitoes" points to the speaker's belief regarding the actual prevalence. **Right:** Hypothetical novice's updated beliefs upon being told that "mosquitoes carry malaria."

closely with those of the speaker) who had general prior knowledge but no experience with the specific situation under discussion. In the language I use in this dissertation, we might say that speakers will endorse a generic statement if it would help align a hypothetical novice listener's understanding with their own. For example, someone who had general prior knowledge but no experience with mosquitoes would know that carrying malaria is an rare trait, based on their experience with the non-mosquito categories and with diseases in general. As such, they would probably assume that, should it turn out that "mosquitoes carry malaria," the generic statement would apply to very few mosquitoes. The hypothetical novice would also probably consider it far more likely that *no* mosquitoes carry malaria based on their experience. These prior expectations are illustrated in the left panel of Figure 1 (inspired by Tessler & Goodman, 2019a). According to this account, speakers endorse "mosquitoes carry malaria" because it would be useful to this hypothetical novice listener, pushing them away from the belief that no mosquitoes carry malaria and towards the belief that a small minority do. This updated belief is illustrated in the right panel of Figure 1.

I return to these semantic accounts in Chapter 5, but for now, the key point is that both accounts suggest that listeners use prior knowledge to make pragmatic inferences about the implied prevalence of a generic statement. Since novices and experts have different prior knowledge by definition, there is the opportunity for miscommunication under either account. For our purposes, these accounts differ in how they would describe the precise way in which differences in prior knowledge would contribute to such a miscommunication. For this dissertation, I remain agnostic about these semantic accounts. I focus instead on comparing how speakers' and listeners' understanding of a generic statement relate to each other, regardless of how they reach their understanding.

I focus on generic statements because the bare noun phrases (i.e., the lack of additional quantifiers, like "sometimes...") provide minimal lexical information and thus place minimal restrictions on the range of plausible interpretations. This flexibility makes generic statements a particularly instructive test case for examining speakers' decisions and listeners' inferences. Bare generalizations are also more fundamental forms of generalization, as indicated by their prevalence in child-directed speech (Pappas & Gelman, 1998; Gelman et al., 1998) and the relative ease with which children can interpret them (Hollander, Gelman, & Star, 2002; Tardif, Gelman, Fu, & Zhu, 2012; Gelman, Leslie, Was, & Koch, 2015). However, there are other quantificational elements (e.g., "often...","mostly...") that similarly leave room for pragmatic interpretation. I suspect experts and novices would exhibit similar patterns in relation to such statements as they do in relation to the statements studied in this dissertation.

Chapter 1

Using Esports as an Experimental Setting for Studying Psycholingustics

Abstract

Esports have become increasingly popular as naturalistic settings for experiments from a variety of disciplines. In large part, this popularity is due to esports helping researchers balance ecological validity and experimental control; esports provide situations in which people are naturally motivated to learn and act in a complex yet restricted environment. Since players often learn and act collaboratively, many researchers have used esports as a setting in which to study communication. However, most of this research has focused on optimizing team performance or player experience, with less work examining fundamental questions of psycholinguistics. The present chapter lays out the unique opportunities esports offers in this regard. Given the focus of this dissertation, I highlight the benefits and limitations of using esports to study psycholinguistics in the context of prior knowledge, emergent expertise, and emergent culture.

1.1 Introduction

A classic joke tells of a farmer whose prized hen stops laying eggs. The farmer seeks out the smartest person in town, a scientist, for advice. The scientist conducts a variety of experiments before she is satisfied that she fully understands the situation. She then goes to the farmer with the good news, saying, "I figured out how to fix your problem, so long as we assume your hen is perfectly spherical and in a vacuum."

While this joke specifically references assumptions made by physicists, it also speaks to the broader tension between experimental control and ecological validity. For experiments to be interpretable, they must be controlled. But we often want to learn things about the real, uncontrolled world. The tension between experimental control and ecological validity arises frequently in psycholinguistics, which examines how language and psychological processes influence and relate to each other. Both phenomena are highly complex, and interactions between the two are even more so. Psycholinguists, like other scientists, often need to sacrifice some ecological validity in order to create well-controlled experiments. For example, if researchers wanted to learn about how people think and talk about a category of birds, a common approach would be to invent a fictional species of bird and give participants limited, carefully controlled information about its features (e.g., Chopra, Tessler, & Goodman, 2019).

Researchers hope that participants' behavior in such experiments yields some insight about real life. But in reality, people think and talk about birds because the topic has some relevance to them. Along the way, they naturally develop varying degrees of expertise about birds, as well as a shared cultural understanding of them. By creating a novel, fully-controlled environment that focuses on an isolated instance of communication and is of minimal significance to participants, researchers risk creating an environment that bears little resemblance to how people think and talk in the real world.

Esports provide an alternative to both traditional experiments and fully naturalistic, real-world settings, providing a unique combination of experimental control and ecological validity. In essence, this approach is a modern adaptation of the line of research that has used games, particularly chess (e.g., Chase & Simon, 1973), as naturalistic experimental settings. Researchers have already begun to use esports as experimental settings for studying communication. For example, game design researchers have used esports environments to investigate which communication mechanisms are effective and satisfying (e.g., Dabbish, Kraut, & Patton, 2012; Wuertz, Bateman, & Tang, 2017). Similarly, industrial/organizational psychologists have used these environments to investigate how high-performing teams communicate (e.g. Richter & Lechner, 2009; Leavitt, Keegan, & Clark, 2016).

In this dissertation, I use esports as a setting in which to study a fundamental question of psycholinguistics. Specifically, I use the game *League of Legends* as a setting in which to study how people use and learn from generic statements, but the benefits of such an experimental setting extend beyond this specific game or research question. In this chapter, I discuss those benefits, while also considering potential limitations.

I first provide a broad overview of how I used *League of Legends* as an experimental setting. I then discuss how my use of this setting points to the potential of esports to provide interpretable yet ecologically valid experimental data. Finally, I consider broader possibilities for using esports as experimental settings for psycholinguistics research in particular and cognitive science research in general.

1.2 Using *League of Legends* to Study Generic Statements

My research questions in this dissertation revolve around how people with different levels of expertise use and interpret generic statements. To study these questions, one would ideally have an experimental setting in which people with different degrees of authentic expertise naturally use generic statements in the context of a naturally-developing culture. *League of Legends* provides just such an experimental setting.

League of Legends is a competitive online game played between two teams of 5 players each. As they play more, players accumulate experience in a complex cultural environment, and they are often highly motivated to do so. There are over 100 million active players (Kollar, 2016), as well as a thriving professional scene. Players are ranked based on their ELO score (Elo, 2008), a system for objectively measuring player performance that was originally developed for chess but which applies to any zero sum game. Throughout this dissertation, participants are classified as experts if they are ranked in at least the 27th percentile of players, indicating that they are well acquainted with the game environment, and they are classified as novices if they report that they have never played the game.

Since *League of Legends* is a team game, players often discuss strategy. They also refine their knowledge of strategy by consulting experts, either via guides and tutorials or via individualized coaching. Discussions of strategy often involve generic statements that describe how a game will likely play out, while crucially allowing for exceptions. A team's primary objective is to destroy the opposing team's base, but there are also numerous secondary objectives that give various forms of assistance towards achieving the primary objective. To prioritize secondary objectives appropriately, players must have some sense of how a given game will tend to progress, while also considering alternative outcomes.

Prior to a game, there is a drafting phase during which each player selects the character they will use. Since the game has not yet begun, players are considering how various characters perform over the course of an entire game—a form of generic statement. And since players are often unfamiliar with their opponents' or teammates' idiosyncrasies, they must be able to make such judgements based on how characters tend to perform when used by a typical player. For example, some characters are generally adept at obtaining secondary objectives that are particularly helpful early in the game, while other characters are generally adept at obtaining objectives that are particularly helpful later in the game. Each game of *League of Legends* is an exchangeable iteration because the environment and characters reset to the same initial state before each game. I can thus measure how broadly a generic statement about a character is being applied in terms of how frequently participants think the generic statement will apply across a set of games.

To summarize the overall methodology in this dissertation, participants are classified as expert or novice based on their reported experience with the game and (in the case of experts) the rank they have achieved within the game. The goal behind this classification is to compare a group of participants who are familiar with the environment under discussion. Within each group, participants are then asked to consider generic statements made about the likely behavior of a character or set of characters. When participants are asked to quantify those generic statements, they are asked to do so in terms of how often a generic statement would apply across a set of 100 games. I provide specific details in the Methods sections of each chapter.

1.3 The Value of Esports as a Research Tool

My use of *League of Legends* as an experimental setting points to how esports can help researchers balance ecological validity with experimental control. Esports provide an environment that is complex and yet has well-defined characteristics. They also provide a setting in which prior knowledge, expertise, and culture emerge naturally.

1.3.1 Complex Yet Well-Defined Environment

One of the main reasons for esports' popularity is that they present rewarding challenges to the people who play and follow them. Yet for all their complexity, the environment is founded on a set of explicit rules specified by the underlying computer code. Games in general are useful settings for studying the impact of prior knowledge because the games' rules help restrict both the learner's goals and the relevant experiences they might have had.

I am far from the first to use game environments to study expertise. Already in the 1970s, researchers were using chess for this purpose (e.g., Chase & Simon, 1973). This research famously showed, for example, that chess masters' remarkable memory for piece positions only applies to sensible game states (i.e., those that have been produced through play rather than randomization). Chess masters are no better than novices at recalling nonsensical details, suggesting that narrative and/or causal reasoning may be relevant to their encoding of those memories. Work on analogous questions in esports has already begun (Bonny & Castaneda, 2016; Lindstedt & Gray, 2019).

In my use of *League of Legends*, I take advantage of a complexity bottleneck. During each game of *League of Legends*, there are infinite permutations as to how players can behave. Prior to each game, however, players must choose which character they will use from a clearly-delineated set of options. In making this choice, players factor in how they expect a given game to play out, with all the complexity that reasoning entails. Furthermore, at last count, there are 170 characters for players to choose from. With one team of five players choosing characters, there are more than one billion possible combinations, although in practice, this number is somewhat reduced by some characters having relatively specialized roles in a team. By setting the experiment in this phase, we can better understand what players are thinking and how they share those ideas with others.

Importantly, this complexity can help with studying the impact of expertise because it leaves plenty of room for differences in prior knowledge. However, the complexity also raises the difficulty of how to go about selecting experimental stimuli from more than a billion options. I chose not to randomly select characters because I wanted to ensure that the stimuli would elicit a full range of responses. In other words, I wanted some situations in which a generic statement would apply quite frequently, some in which it would not, and some that were borderline. Also, not all of the one billion possible combinations of characters are equally likely or even equally sensible choices. Fortunately, there are publicly-available statistics about the rate at which each character is chosen. I used these statistics to select sensible experimental stimuli more efficiently.

Another benefit of studying generic statements in this environment is that it helps us delimit the topic of the relevant generic statements. In the real world, it can be difficult to know what entity people are referring to when they use generic statements. Statements like "mosquitoes carry malaria" refer to mosquitoes as a general concept rather than a specific set of mosquitoes. But what is the category of things in the world that this concept picks out? There are many species of mosquitoes, some of which carry malaria and some of which do not. Within those species, every individual is different. When we say "mosquitoes," we refer to some idealized essence of "mosquito-ness" that may not actually exist (Gelman, 2003). In contrast, the characters in *League of Legends* get reset to the same initial state before each game. When discussing a specific characters in that initial state, we know precisely what entity we are talking about, because the characters are defined by properties assigned in the underlying computer code.

Using *League of Legends* as an experimental setting helps strike a balance complexity and experimental control. The stimuli are sufficiently complex to provide authentic examples of generic statements and associated higher-level reasoning. The environment is sufficiently well-defined to ensure that the stimuli are ecologically valid and that we understand both the strategic context (i.e., available choices) in which the generic statements are being considered and the entities being referenced by those generic statements.

1.3.2 Studying Prior Knowledge and Expertise Using Esports

Esports offer unique opportunities to study expertise. The standard approach to studying expertise in the lab is to create a simplified approximation, allowing researchers to manipulate or control participants' expertise. Researchers can then precisely quantify expertise, but they must also commit to a specific definition of expertise. For example, one common approximation of expertise is for researchers to vary the amount of information given to participants (Hawthorne-Madell & Goodman, 2019). With such a procedure, researchers implicitly assume that expertise is a function of the quantity of experiences upon which one's prior knowledge is based. In committing to a particular definition of expertise, researchers risk creating an invalid approximation.

Prior knowledge and expertise are related concepts, and indeed the two are often used interchangeably because experts can simply be defined as those who have accumulated more prior knowledge (e.g., Helton, 2007). Yet experts are also expected to be particularly effective at using their prior knowledge, which is not necessarily the same thing as having a lot of it. Expertise is thus often characterized by a fundamental shift in how prior knowledge is organized (Chi, Feltovich, & Glaser, 1981). Moreover, expertise can be thought of as a social phenomenon (Mieg & Evetts, 2018): someone is an expert if others recognize their authority as an expert.

Choosing how to define expertise is less problematic in a naturalistic experimental setting because such a setting guarantees that the underlying expertise is at least authentic to that setting. However, researchers do still need to commit to a metric for expertise. How would we go about establishing whether someone is a mosquito expert? If we think of expertise as a social phenomenon, an appropriate operationalization might be whether others have conferred credentials upon a participant, such as prestigious degrees or publications. The credentials we identify as relevant would be largely subjective. Is it more important for an expert to have a Ph.D. or numerous publications? If expertise is instead considered a cognitive phenomenon, it might be more appropriate to test knowledge or skill. In this case, the contents of the test would often be subjectively selected. Is it more important for an expert to be able to identify different species of mosquitoes or to explain their physiology? The point is that there are many approaches to defining expertise, making it difficult for researchers to arrive at a satisfactory operationalization.

Esports provide situations in which expertise develops naturally and yet researchers have a reasonable chance of measuring it objectively. As such, considerable research has already been conducted looking at how players become experts (Röhlcke, Bäcklund, Sörman, & Jonsson, 2018; Bonny & Castaneda, 2016), how their expertise impacts the way they engage with a given game (Reeves, Brown, & Laurier, 2009; Bonny & Castaneda, 2016), and whether the skills they develop transfer to other domains (Green & Bavelier, 2006; Seya & Shinoda, 2016). Players are naturally self-motivated to find information that is relevant to their goals and ignore that which is not. In the process, they have the time and motivation to reorganize and grow their knowledge. They also interact socially with other players who have varying levels of expertise. As players become more accomplished within the environment, there is typically a metric tracking their progress, making it relatively simple to measure their expertise. While such a metric is not perfect, given that it exclusively measures performance rather than knowledge or social authority, it is far more convenient than the approximations (e.g., tests, surveys, observations, interviews) that would typically be available for measuring naturally-emerging expertise.

My experimental questions focus on how familiarity with a domain, or lack thereof, relates to how people use and interpret generic statements about that domain. It is thus crucial to establish whether a given participant is familiar with *League of Legends*. We could rely on common approximations, such as asking participants to demonstrate their knowledge on a test or report the extent of their experience, but *League of Legends* provides a more convenient option in the form of a metric for players skill. As with many zero-sum games, this metric is based on ELO rating (Elo, 2008), which is an objective measure of performance originally created for chess. Players' ELO rating increases if they win and decreases if they lose, with bigger changes after surprising results (i.e., winning against a strong opponent or losing against a weak one). In the case of *League of Legends*, ELO is used to divide players into different tiers, which is how expertise is measured in this dissertation.

By using these tiers as a metric for expertise, expertise is defined in terms of proficiency; you are an expert if you are good at the game. As discussed above, this is far from the only way to define expertise. Players can know a lot about the game without being particularly skillful or having much social authority in the community. However, these three considerations are presumably correlated in most situations. Higher-ranked players are more likely to know more about the game and to have more authority when talking about it. And while players may have their own notions of how to have fun, being good at the game is surely a central goal.

This operational definition also implicitly assumes that higher-ranked players know more about the game than lower-ranked players. For subtle distinctions in rank, such an assumption may be problematic, but to address the present research questions, we simply need to divide participants into two groups based on whether or not they have directly-applicable prior knowledge. In this dissertation, I assume that players who have attained the rank of Silver-level (27th percentile; Milella, 2021) or higher do have directly-applicable prior knowledge, whereas participants who have no prior experience with the game do not have directly-applicable prior knowledge. Since we can safely assume that players ranked in the Silver-tier or higher are well-acquainted with League of Legends, this operational definition is satisfactory for my purposes.

Esports can also help with studying the hierarchical nature of prior knowledge (Kemp, Perfors, & Tenenbaum, 2007), which I consider in the "Discussion" section of Chapter 2 and in the "Developmental Trajectory of Generealizations" section of Chapter 5. While novices have no experience with this specific game, they do at least have prior expectations about games in general. Experts have this very general knowledge about games as well, but within that knowledge of games, they also have a sense of how this specific game works, of how characters within the game work, and maybe even how a specific combination of characters works.

Esports offer unique opportunities to study hierarchical prior knowledge, which I have yet to fully exploit by focusing exclusively on *League of Legends*. In particular, there are numerous examples of games which share their general structure but differ in their specific feature. For example, *Defense of the Ancients* (*DOTA*), is identical to *League of Legends* in its basic structure but different in its specifics. By recruiting experienced *DOTA* players for a *League of Legends* study, researchers could tease apart the effects of general prior knowledge (common to both *DOTA* and *League of Legends*) from those of game-specific knowledge. Given the landscape of video games, this opportunity is not unique to *League of Legends*. Sequels create games with similar general mechanics but different specifics. Games are also readily divided into genres as developers borrow mechanics from each other. In all of these situations, players with experience in similar games already have general prior knowledge that may help inform and accelerate their acquisition of more specific prior knowledge.

1.3.3 Studying Emergent Culture Using Esports

Esports also provide situations in which players are communicating in the context of a naturally-developing culture. That culture is focused on learning about an environment with well-defined characteristics and a unified goal, making it simpler to ensure that experimental materials are appropriately situated within that culture. The environment in esports restricts the decisions that players can make in response to cultural factors, and it also provides a standardized reward structure. Researchers have already begun to use esports as microcosms in which to observe a variety of cultural phenomena, such as stereotype threat (Kaye, Pennington, & McCann, 2018) and social systems for regulating behavior (Kou & Nardi, 2013, 2014). Numerous studies have also examined the cultural dynamics of cooperative play, both in terms of stable teams developing a system of shared meaning specific to themselves (e.g., Richter & Lechner, 2009; Dabbish et al., 2012) and the formation of ad-hoc teams in the context of the general culture of the game (e.g., Kou & Gui, 2014; Kim, Keegan, Park, & Oh, 2016).

In developing the experiments for this dissertation, I wanted to make sure that the communication scenarios are properly situated in an authentic culture. *League of Legends* provides such a culture because players often work collaboratively and can tap into a wealth of pre-existing knowledge accumulated by the larger gaming community. In the process, the community naturally develops a shared system of meaning. Fortunately, this culture is relatively narrow in scope since players are focused on a game environment, which is itself restricted.

I made sure that each statement is representative of how players communicate in this cultural context by situating the statements in the time between games. Without a specific match to reference, players discuss the strengths and weaknesses of various characters in an abstract sense. Situational variables, such as the relative skill of opponents, get subsumed within the generic statements. For example, a character might generally do well in the early game, but skilled opponents might be able to prevent this from happening. Such a consideration is part of why the statements found in the stimuli are generic statements rather than a deterministic statement of fact. If players do not have a specific opponent in mind, the potential impact of relative skill gets factored into their thought process about generic statements as a source of exceptions to the general trend.

To simplify the experiments, I used predetermined statements. Such a setup is still realistic because players often learn from non-interactive sources like forum posts and expert guides. However, it does artificially restrict the choices available to participants. In principle, I could study open-ended and dynamic interactions, particularly since the structure of *League of Legends* would be conducive to such an experimental design. As with other team-based computer games, players often speak to each other. In *League of Legends*, there is a phase before each game in which players are deciding which characters to select. During this phase, players need to align their expectations about how characters are likely to perform during the forthcoming game, presenting an excellent opportunity to observe generic statements in action.

League of Legends is far from the only esport that provides convenient opportunities for studying communication in a naturalistic cultural context. Some games provide players with a set of streamlined communication options, which allows researchers to observe how players use a communication system that is already controlled (e.g., "pings"; Wuertz et al., 2017). Players will also sometimes communicate verbally, allowing researchers to observe players' use of language in a controlled context–typically a single audio channel where people discuss a shared objective reality, which happens to be defined and recorded by a computer.

1.3.4 Limitations

However, there are some limitations on using esports as experimental settings for psycholinguistics. First, while I argue that esports strike a nice balance between ecological validity and experimental control, any attempt to make an experiment more ecologically valid introduces the possibility of increased noise relative to a controlled experimental setting, and esports are no exception. Second, what participants are asked to do needs to make sense in the context of the structure and culture of the game. Unlike laboratory-controlled experiments, experimenters cannot design a situation to their exact specifications. Instead, experimenters must work with what is already provided by each game, carving out procedures and stimuli from what would happen naturally.

There are also limitations in generalizing results to a broader population. Most fundamentally, players are self-selecting. People decide whether, and how seriously, they want to play, so there may be consistent differences between players and non-players beyond experience with the game. Comparisons between existing players and non-players or between experts and non-experts are only correlational. Experimenters can recruit non-players and randomly assign some to undergo training (e.g., Richter & Lechner, 2009), but such an approach is labor-intensive. Furthermore, findings from such an experiment may not generalize to people who would choose to play the game on their own and are therefore ineligible for such a study.

On the bright side, my results indicate that experienced players and undergraduates have similar demographics, at least along some dimensions (i.e., age, education, languages spoken). However, the sample of experienced players was heavily male-dominated. Such a disparity is unfortunately symptomatic of a lack of gender diversity in the esports community at large (Kuznekoff & Rose, 2013; Ruvalcaba, Shulze, Kim, Berzenski, & Otten, 2018). This situation may improve, but, in the meantime, researchers must be cognizant of potential gender disparities when extending findings to the broader population.

Linguistically, there is no reason to expect gender to influence the results presented in this dissertation, but gender could interact with game-specific knowledge in a number of ways. Male and female players develop their gaming skills—and presumably their knowledge about the game—at the same rate (Ratan, Taylor, Hogan, Kennedy, & Williams, 2015), but female players disproportionately use female characters and may thus have expertise about different aspects of the game than males do (Ratan, Fordham, Leith, & Williams, 2019). Gender may also influence social aspects of both expertise and communication, particularly in the context of persistent gender discrimination (Kuznekoff & Rose, 2013; Ruvalcaba et al., 2018). While these experiments can be replicated with gender-matched samples, it is a fundamental limitation of the esports setting that the gender of participants is so skewed.

1.4 Closing Thoughts

The features which make *League of Legends* so suitable for studying my specific research questions also apply to a variety of esports and a variety of related questions. Esports provide environments in which people are highly motivated to think and talk. The environments are rich, and yet they are self-contained and have well-defined characteristics, opening up exciting possibilities for designing experiments. As with any experimental setting, esports have limitations. However, they provide unusual opportunities and shortcuts for balancing ecological validity and experimental control. As esports become increasingly popular, they offer researchers a way to study how players communicate to each other within and about the games. In doing so, they present opportunities to (re)examine fundamental questions in psycholinguistics.

Chapter 2

Experts and Novices Interpret Generic Statements Differently

Abstract

Generic statements can vary widely in terms of their implied prevalence—the specific quantification they convey. Recent research suggests that listeners use their prior knowledge to interpret how broadly generic statements apply (Tessler & Goodman, 2019b). Since experts and novices have different prior knowledge, they may differ in how they interpret generic statements. In this chapter, I describe a study that compares how experts and novices interpret generic statements. As hypothesized, experts interpreted generic statements more flexibly than novices did, and novices tended to assume generic statements applied more broadly than experts did.

2.1 Introduction

We begin this investigation of how generic statements function as pedagogical tools by first considering how listeners interpret generic statements. More specifically, this chapter describes a study examining how listeners interpret the implied prevalence of a generic statement. This study was conducted in collaboration with Alexander Etz, Gregory Scontras, and Barbara Sarnecka,

The first research to directly examine this question found that listeners tend to interpret generic statements rigidly and broadly (Cimpian, Brandone, & Gelman, 2010). In their series of experiments,"listener" participants were told that a speaker had decided to make a generic statement about a novel animal category and were asked to interpret that generic statement in terms of how broadly it applied (e.g., "A person who is familiar with luzeks decided to say that 'luzeks have purple tail feathers.' What percentage of luzeks do you think have purple tail feathers?"). Participants consistently interpreted generic statements about the physical features of a novel animal category as implying near-universal (≈90%) prevalence.

In a recent experiment, Tessler and Goodman (2019b) showed that listeners do not always interpret generic statements so rigidly. In interpreting a generic statement, listeners have two sources of information: the generic statement itself and their own prior knowledge, which can temper any tendency toward assuming near-universal prevalence. Applying this idea to the mosquito example, a listener will presumably distinguish between the how broadly the statements "mosquitoes carry malaria" and "mosquitoes are insects" apply if they are familiar with mosquitoes, insects, and malaria. So long as listeners have prior knowledge comparable to that of a speaker, they should be able to adjust their interpretations according to the context of a statement resulting in interpretations that are comparable to what the speaker had in mind.

But although generic statements are sometimes interpreted by listeners with prior knowledge comparable to that of the speaker, they are also used to help less-experienced listeners learn. By definition, novices and experts have different prior knowledge, which may lead them to interpret the same generic statements differently. In the case of Cimpian et al. (2010)'s experiment, listeners were trying to use generic statements to learn about a category with which they had no prior knowledge. In contrast, Tessler and Goodman (2019b) provided listeners with prior knowledge, and indeed manipulated that prior knowledge. Thus listeners' ability to interpret generic statements flexibly may depend on their prior knowledge.

In the following study, my collaborators and I examine whether experts and novices differ in how they interpret generic statements. As discussed in Chapter 1, we use the esport *League of Legends* as the experimental setting. We measure participants' interpretation of the implied strength of a generic statement in terms of how frequently they expect it to be exemplified in a set of iterative games. Our key research questions are whether (1) experts will have flexible interpretations that take into account prior knowledge, (2) novices will have less flexible interpretations, and (3) novices will tend to interpret generic statements as more broadly applicable than experts do.

2.2 Methods

2.2.1 Participants

Expert participants ($n = 49$) were recruited from *League of Legends* online forums. They needed to be ranked in the 27$^{\text{th}}$ percentile (Silver tier) of players or higher to ensure that they were well acquainted with the domain. Novice participants ($n = 33$) were recruited from UC Irvine's undergraduate participant pool. Participants who indicated prior familiarity with the game were removed from the novice group, and those who qualified as expert participants were reclassified as such. Novice participants were also removed from the dataset if they

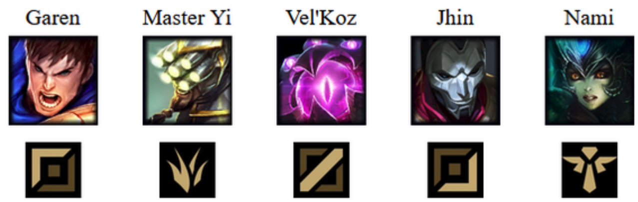

Figure 2.1: Team composition example. Each icon indicates the specific team role being fulfilled by that character.

spent less than 15 seconds reading the background information provided about *League of Legends* (of the original sample of 53 participants, 17 were eliminated for this reason).

While these samples were very similar in terms of education and first language, there were considerable gender disparities. The sample of expert participants was heavily male-dominated (89% male), whereas the sample of novice participants was heavily female-dominated (91% female). Again, novice participants were undergraduates who were recruited at random from UC Irvine's undergraduate participant pool. While we removed novice participants who indicated that they had prior experience with the game, this criterion only resulted in the removal of 2 females and 1 male.

2.2.2 Materials

Participants completed an online survey in which they interpreted what a hypothetical expert player (henceforth the "knowledgeable speaker") is trying to communicate by endorsing a generic statement.

These generic statements pertain to a team's *composition*—the set of 5 characters that make up a team, as illustrated in Figure 2.1. Specifically, the generic statements are about whether a given composition excels in the early game, meaning that the players should try to build an early advantage, or in the late game, meaning that the players should try to stall the game until they are able to realize their potential. Importantly, while the generic statements can

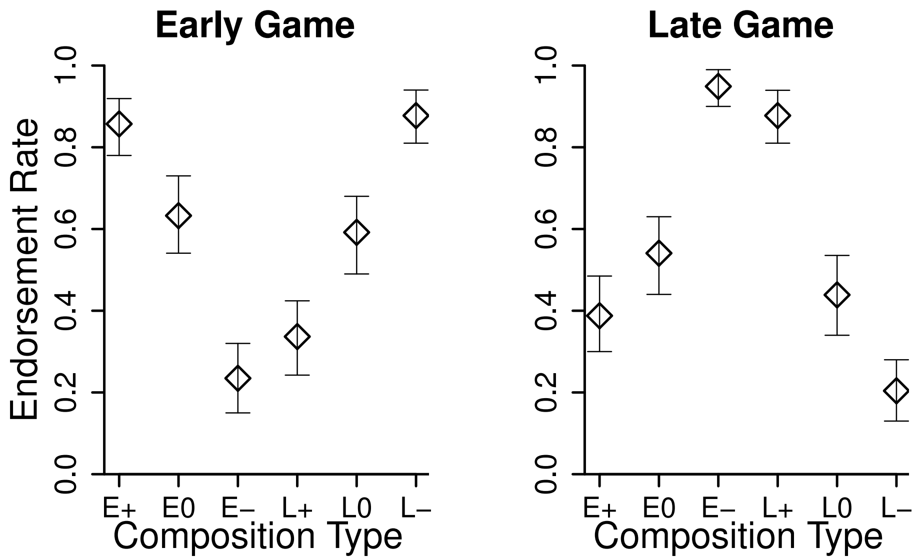

Figure 2.2: Proportion of experts who would endorse the generic statement that the given composition type excels in the early game (left) or late game (right). Error bars are 95% bootstrapped confidence interval.

express trends, experts should know that contextual factors will impact how things proceed in any given game.

Novice participants were told that compositions could vary in terms of how likely they are to excel in the early and late game. While they were not explicitly told about the extent to which compositions can be the victim of circumstances, our results imply that novice participants inferred a probabilistic causal link between team composition and the ability to excel in the [early/late] game, rather than a deterministic one.

The task began with some background information. Participants saw broad explanations of how the game works, team compositions, and the phases of a game (i.e., early and late).

Following the background information, participants saw a possible composition (see Figure 2.1 for an example). Participants were first asked whether, in order to explain to a friend how this composition works, they would themselves endorse either of two generic statements (by choosing “yes” or “no” to indicate whether they would say “this team excels in the [early/late] game"). These questions confirmed that the generic statements in this study were ones that real players would make (see Figure 2.1). Novices had no basis for making these judgements, so their responses do not vary by composition type.

Participants were then asked how frequently they would expect the specific team composition to excel in the [early/late] game across a subset of 100 games (by adjusting a slider with endpoints of 0 and 100). While this question is not central to the current study, we do use it to evaluate competing interpretations.

Next, participants were asked to imagine that, having seen the composition play 100 games against varied opposition, the knowledgeable speaker endorsed the generic statement ("this team excels in the [early/late] game"). Participants were then asked to interpret the intended strength of the knowledgeable speaker's generic statement in terms of what they thought had happened in the subset of games on which the knowledgeable speaker's decision was based. In other words, given that the knowledgeable speaker made the generic statement, in how many of the 100 games seen by the knowledgeable speaker do participants think the composition excelled in the [early/late] game? Participants adjusted a slider with endpoints labeled 0 and 100 to indicate their response.

Finally, participants were asked to describe their experience with League of Legends in terms of time played, ELO ranking (Elo, 2008), and knowledge of the game's characters. This information was used to verify the classification of participants as either novices or experts.

2.2.3 Design

The present study can be summarized as a 2 (expertise) x 6 (composition type) design. Based on pilot data, our stimuli included 12 team compositions that varied in terms of how often experts think they would excel in the early or late game.[1]

Specifically, we selected two compositions for each of six types: strong early game, middling early game, weak early game, strong late game, middling late game, and weak late game (abbreviated as E_+, E_0, E_-, L_+, L_0, and L_-). Participants thus considered a total of 12 compositions.

Although each of the 12 compositions was selected with only their early- or late-game ability in mind, a composition that excels in the early game is generally less likely to excel in the late game and vice versa. This assumption is supported by the proportion of expert participants who would endorse the generic statements themselves (see Figure 2.2).

Participants considered 6 generic statements that the knowledgeable speaker did endorse and 6 generic statements that the knowledgeable speakers did not endorse (randomly assigned and counterbalanced across composition type). In the present study, we only consider participants' interpretations when the knowledgeable speaker *did* endorse a generic statement, given that we are interested in how participants interpret a generic statement, not their reasoning about why a speaker chose not to make a generic statement.

We did not vary the two generic statements ("This composition excels in the [early/late] game.") by composition type; experts were occasionally asked to interpret generic statements with which many of them would disagree. In fact, such disagreement happened 42.7% of the time.

[1] Candidate compositions were randomly generated at the rate at which they are used in the game and evaluated by at least 10 experts.

2.3 Results

2.3.1 Modeling Approach

Our dependent variable was participants' estimates of how broadly a generic statement applies. We recorded participants' responses as proportions, which represented participants' stated beliefs about how often a given composition had excelled in the [early/late] game across the subset of games on which a given generic statement was based.

To analyze the results, we developed a mixed-effects linear model. For the ith participant viewing the jth composition, we modeled their response, y_{ij}, as the additive result of three influences: an expected response for the given condition, a participant-level offset, and trial-to-trial noise. The expected response in the context of a given composition type is represented by a β parameter. Novices have a single β_0 given their lack of domain-specific prior knowledge that would allow them to modulate their expectations by composition type. In contrast, experts have a β_j specific to each of the j composition types.

The ith participant's offset, α_i, is modeled as a random intercept being drawn independently from a normal distribution with a mean of zero and standard deviation τ. A positive value for α_i indicates that the participant tends to provide high responses compared to other participants, with negative α_i indicating the reverse. Finally, independent noise present on a given trial, ϵ_{ij}, is represented as being independently drawn from a normal distribution with a mean of zero and standard deviation σ.

Thus, the model can be written as

$$y_{ij} = \beta_0 x_0 + \sum_{j=1}^{6} \beta_j x_j + \alpha_i + \epsilon_{ij}, \tag{2.1}$$

with $\alpha_i \sim N(0, \tau^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$. Each x is a binary indicator of whether the influence represented by the associated β is relevant for participant i looking at a given composition of type j. For novices, $x_0 = 1$, with all other $x_j = 0$; in this case, the model reduces to

$$y_{ij} = \beta_0 + \alpha_i + \epsilon_{ij}, \tag{2.2}$$

because novices presumably do not distinguish between composition types and thus do not vary their response by composition type. We examine the validity of this assumption later.

The x_1 through x_6 variables indicate, for experts, the composition type for a given trial; only one x_j will ever equal 1 on a given trial. For a trial where an expert is viewing composition type j, the model reduces to

$$y_{ij} = \beta_j + \alpha_i + \epsilon_{ij}. \tag{2.3}$$

This same model was applied separately to early game responses vs. late game responses, as it is not obvious *a priori* that these two generic statements are equivalent.

This modeling approach allows for inference on both the group-level (βs) and participant-level (α_is). Each participant only saw six generic statements each, so any inferences about an individual α_i are necessarily limited. However, including α_is in the models accounts for potential dependency in a participant's repeated responses and helps to ensure an appropriate level of uncertainty in the estimation of the βs.

The models were implemented in JAGS (Plummer, 2003) and fit to the data using Bayesian parameter estimation. Each β was given an independent $N(0.5, 0.25^2)$ prior distribution, with almost all of its mass between 0 and 1. The mean of 0.5 was chosen because, in the context of this experiment, 0.5 would be a "neutral" response (i.e., the composition exemplified the referenced trait—excelling in the [early/late] game—as often as not). The variation of the participant-level offset, τ, and the trial-to-trial noise, σ, were both given a gamma prior with

Table 2.1: Inferred descriptive statistics by composition type (β_0 indicates novices). Approximate credible intervals for a β can be formed by taking the mean plus or minus 2 times the SD.

Interpretation of Early Game Generic statements							
β	E_+	E_0	E_-	L_+	L_0	L_-	0
Mean	.65	.62	.56	.57	.60	.66	.70
SD	.02	.02	.02	.02	.02	.02	.02

Interpretation of Late Game Generic statements							
β	E_+	E_0	E_-	L_+	L_0	L_-	0
Mean	.59	.63	.71	.68	.58	.57	.69
SD	.03	.03	.03	.03	.03	.03	.02

shape 1.5 and rate 2. The shape and rate parameters were chosen to not give too much credence to unreasonably large values of τ and σ.

2.3.2 Modeling Results: Estimation and Fit

As shown in Figure 2.3, the models fit the data well, with inferences closely matching the empirical means. There is a U-shaped pattern in the empirical responses for the experts, depending on composition type and the content of the generic statement (early vs. late). In contrast, the empirical responses of novices show no clear pattern for the models to match, but their mean responses by composition type are still close to the model's β_0 estimate. A summary of the posterior distributions for the βs can be found in Table 2.1.

Variability in responses due to systematic differences between participants (τ) was comparable to variability due to noise (σ). For generic statements about the early game, the posterior estimates of σ and τ are 0.11 (95% CI [0.10, 0.119]) and 0.12 (CI [0.10, 0.14]) respectively. For generic statements about the late game, the posterior estimates of σ and τ are 0.13 (CI [0.12, 0.14]) and 0.13 (CI [0.11, 0.15]). These magnitudes are directly comparable, as they are on the same scale, but can be more conveniently interpreted once they are converted into the familiar correlation coefficient.

For a mixed-effects model using random intercepts, the implied correlation between two responses given by a participant is $\rho = \tau^2/(\tau^2 + \sigma^2)$; note that the correlation between any responses from different participants is zero by assumption, after accounting for composition type and expertise. The value of ρ indicates how predictable a participant's responses will tend to be. Performing this transformation using the relevant posterior samples, we obtain posterior estimates for ρ of 0.53 (CI [0.46, 0.63]) and 0.50 (CI [0.42, 0.60]) for the early game and late game generic statements, respectively. These correlations indicate substantial individual differences across participants.

2.3.3 Modeling Results: Hypothesis Testing

Question 1

Our first research question was whether experts would interpret generic statements flexibly, varying how broadly applicable they think a generic statement is depending on the context in which the generic statement was made. More specifically, we might expect experts to interpret generic statements as applying more narrowly when their own prior knowledge conflicted with the generic statement, and more broadly when the generic statement agreed with their prior knowledge. Figure 2.2 shows how often experts were willing to endorse a given generic statement themselves, indicating whether their prior knowledge conflicts with it. For trials in which the generic statement referred to a composition's performance in the early game, experts would interpret generic statements about composition type 1 (strong early game) as more broadly applicable than those about composition type 2 (middling early game), and would in turn interpret generic statements about composition type 2 as more broadly applicable than those about composition type 3 (weak early game). We expected this pattern to be inverted for the late game composition types. Furthermore, we would expect these two sets of orderings to be reversed when generic statements referenced the late game. This hypothesis translates into the following simultaneous orderings: $\beta_{E_+} > \beta_{E_0} > \beta_{E_-}$ and $\beta_{L_+} < \beta_{L_0} < \beta_{L_-}$ for early-game generic statements, and $\beta_{E_+} < \beta_{E_0} < \beta_{E_-}$ and $\beta_{L_+} > \beta_{L_0} > \beta_{L_-}$ for late-game generic statements.

As shown in Figure 2.3, the hypothesized orderings are clearly reflected in the early-game model posteriors and slightly less clearly in the late-game model posteriors. We quantify support for this hypothesis by computing Bayes Factors, comparing the posterior odds for the orderings to the prior odds. Complex hypotheses involving many simultaneous order constraints necessarily have very low prior odds because there are many ways *a priori* for the ordering to be violated. There are 36 ways to simultaneously order two sets of three parameters. Thus, the two hypothesized orderings each have prior odds of 1:35. We calculated the posterior odds of the hypothesized orderings by finding a) the number of joint posterior samples simultaneously obeying the order constraints, finding b) the number of samples violating the constraints, and dividing a by b.

The hypothesized orderings have posterior odds of 8.1:1 in the early-game model and posterior odds of 1.3:1 in the late-game model. Comparing the posterior odds to the prior odds, we obtain Bayes factors for the early game and late game hypotheses of about 280 and 46, respectively. Even in the case of the late-game model where uncertainty remains regarding the ordering of β_5 and β_6, the data provide relatively strong evidence in support of the hypothesized orderings.

Question 2

Our second research question was whether novices would interpret generic statements rigidly (i.e., consistently), responding similarly regardless of composition type. In fact, an assumption of our models only allows for one answer to this question, since only β_0 is available to dictate the expected novice response. This assumption is informally supported by the close correspondence between novices' empirical means and the models' inferred estimates (see Figure 2.3). To test this assumption formally, we created alternative models in which expected responses for novices could vary in response to composition type. The priors chosen were the same as those used for the experts' parameters. Model comparisons were done using the Deviance Information Criterion (DIC), a metric that accounts for the inherent trade-off between model fit and complexity (Spiegelhalter, Best, Carlin, & van der Linde, 2014). The expanded models do little to improve our account of the data (ΔDICs of 8 and 7 favoring the simpler early and late models, respectively), evidence suggesting that novice interpretations are unaffected by composition type.

The above test only considers one aspect of rigidity, namely that the context makes no difference to how novices interpret the generic statements. Another aspect of rigidity is the relative consistency of novice and expert responses on a trial-to-trial basis. In other words, are novices' responses truly more rigid, or does their variability come in the form of general noise? Whereas composition irrelevance is reflected in the β parameters, trial-to-trial consistency is captured by the distribution of the σ noise parameter.

To test whether trial-to-trial noise differed between novices and experts, we expanded the model so that each ϵ_{ij} is drawn from one of two normal distributions, with different variances depending on the expertise of a given participant. The priors for the new parameters, which we call σ_{exp} and σ_{nov}, were the same as that of σ in the original model.

For early game generic statements, the posterior estimate of $\sigma_{exp} - \sigma_{nov}$ is -.01 (CI [-.02, .01]). For late game generic statements, the posterior estimate of $\sigma_{exp} - \sigma_{nov}$ is -.02 (CI [-.04, 0.00]). These estimates suggest that any differences that may exist between σ_{exp} and σ_{nov} would be relatively minor. Indeed, a null hypothesis test using the Savage-Dickey method (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010) provides evidence for the equivalence of σ_{exp} and σ_{nov} (early game and late game BF_{01} of about 60 and 5.5, respectively).

On the face of it, this result indicates that experts and novices exhibit similar trial-to-trial noise, and suggests that novices do indeed interpret generic statements rigidly when lacking domain-specific prior knowledge. However, an exploratory model, which we describe in Section 2.4 "Two Types of Expert Disagreement?", brings this finding into question. Instead, it appears that novices are not necessarily rigid in their interpretations, but rather unsystematic in their flexibility.

Question 3

Our third research question was whether novices would give higher responses on average. Figure 2.3 shows that context dictates how experts respond. In some contexts, the average responses for experts are just as high as those of novices. When the groups do diverge, it is clearly because experts' responses are lower, indicating that experts are interpreting some generic statements more narrowly.

2.3.4 Two Types of Expert Disagreement?

We also considered the possibility that experts might exhibit two distinct types of behavior when interpreting generic statements that they would not make themselves. As illustrated in Figure 2.2, participants disagree with the speaker fairly frequently in this experiment, particularly since some stimuli were specifically selected as being relatively poor examples of a generic statement.

Participants were instructed to assume that the knowledgeable speaker was trustworthy. We thus expected expert participants to interpret the knowledgeable speaker saying something unexpected as indicating that the knowledgeable speaker had seen something unexpected, namely something that would justify the generic statement. By this reasoning, we would expect participants to interpret the generic statement with which they disagree in a manner similar to how they would interpret generic statements with which they agree. We can model this behavior using the same approach we previously applied to all interpretations.

However, expert participants may occasionally ignore the instruction about trusting the knowledgeable speaker, or assume that the speaker had actually seen the "truth" as the expert participant understands it and is misremembering or misinterpreting what they saw in making the generic statement. In both situations, we would expect expert participants to simply reproduce their own prior expectations when interpreting a generic statement with which they disagree. This alternative approach is only available to experts, since novices lack the prior knowledge required to disagree with the knowledgeable speaker.

Our concern in considering this second form of disagreement is that, if experts are recapitulating their own prior expectations independent of what the knowledgeable speaker said, they might not be truly "interpreting" the generic statements. This point is debatable, but for now our goal is to examine the extent to which this form of behavior impacts our findings.

We incorporate the two types of disagreement into our model by adding a latent mixture component. For novices and for generic statements about which an expert agrees with the knowledgeable speaker, the first component of the model applies the structure described previously. When experts disagree with the generic statement, the model tries to determine if they are interpreting the generic statement as we expect them to or if they are simply returning their prior beliefs as their response.

In this expanded model, we use the fact that we have measured each expert's prior expectations for every trial. The model attempts to classify each response on disagreement trials as arising from one of two processes: $\psi = 0$ if the expert is incorporating the speaker's statement, or $\psi = 1$ if they are returning their prior expectation. This full model for experts can be specified as

$$y_{ij} = \begin{cases} \beta_j + \alpha_i + \epsilon_{ij} & \text{if } \psi_{ij} = 0, \\ x_{ij} + \epsilon'_{ij} & \text{if } \psi_{ij} = 1. \end{cases} \tag{2.4}$$

where x_{ij} represents the ith expert participants' prior expectations about how broadly the jth generic statement would apply. We model ψ_{ij} as a draw from a Bernoulli distribution with weight ϕ_i, representing the ith expert's tendency to respond using only their prior expectation. For example, if $\phi_i = .80$, the ith participant returns their own prior expectations as their response 80% of the time when disagreeing with a generic statement. Each individual's ϕ_i probabilities are modeled as being drawn from a truncated normal distribution with mean μ_ϕ and standard deviation σ_ϕ. Thus μ_ϕ represents our experts' group-level tendency towards using their own prior expectations when responding to a generic statement with which they disagree, and σ_ϕ represents how consistent our experts are in this regard.

For the first component of the model (when $\psi_{ij} = 0$), we use the same priors as those described in the above *Modeling Approach* section. For the second component's parameters, we assign μ_ϕ a uniform prior between 0 and 1, and we assign σ_ϕ a gamma prior with shape 4 and rate 10. We also specify a unique trial-by-trial noise parameter (ϵ'_{ij}) for when participants are returning their prior expectations. We give this parameter a more restrictive gamma prior, with a shape of 0.5 and a rate of 8, reflecting our assumption that recapitulating one's own prior expectations is likely to be a less noisy process than inferring the knowledgeable speaker's beliefs.

Model comparison indicates that the two-process latent mixture model performs much better than the single process model (ΔDICs of 188 and 106 in favor of the early and late latent mixture models, respectively). We visualize the results of our latent mixture model in

Figure 2.4. The graphs are divided based on whether the generic statement the participant disagreed with referenced early game (top two rows) or late game proficiency (bottom two rows). The graphs are further divided based on the composition referenced in the generic statement (E+, E0, E-, L+, L0, or L-). The number of dots in each graph indicates how often participants disagreed with each generic statement. For example, few experts disagreed with the generic statement that the E+ composition "excels in the early game" (Early E+) because that composition was specifically selected as being a good example of one which excels in the early game. The fill color of each dot represents the model's confidence in that trial having arisen from one process or the other. White dots indicate that the response is very likely an example of the expert listener recapitulating their own prior expectations rather than assuming the knowledgeable speaker saw something unexpected. Our posterior estimate for the group-level μ_ϕ parameter is 0.159 (95% CI [0.0051, 0.312]), indicating that there is a small but potentially noteworthy minority of trials on which experts disagree with the knowledgeable speaker by using their own prior expectations as their interpretation of the generic statement. Looking at the overall number of white dots across the graphs supports this interpretation. Our posterior estimate for σ_ϕ is 0.349 (95% CI [0.171, 0.640]), indicating considerable individual differences in the extent to which participants use this response strategy.

To assess the impact of such trials on our original findings, we conducted the same hypothesis tests.[2] As we would expect, the U shapes created by the β_j parameters describing experts' interpretations became shallower. Experts who recapitulate their own expectations when interpreting generic statements they disagree with tend to have lower responses than those who interpret such generic statements as indicating that the knowledgeable speaker saw something unexpected. Even so, we find similar evidence that experts are following the ordering constraints described in *Question 1* above (BF_{10} of 80 and 125 for early and late game generic statements respectively; previously 280 and 46). Thus experts are still

[2]For specifics, please refer to supplementary materials available at (https://osf.io/vae2b/)

Table 2.2: Inferred descriptive statistics by composition type (β_0 indicates novices) under latent mixture model. Approximate credible intervals for a β can be formed by taking the mean plus or minus 2 times the SD.

Interpretation of Early Game Generic Statements							
β	E_+	E_0	E_-	L_+	L_0	L_-	0
Mean	.67	.66	.60	.59	.64	.68	.70
SD	.02	.02	.02	.02	.02	.02	.02

Interpretation of Late Game Generic Statements							
β	E_+	E_0	E_-	L_+	L_0	L_-	0
Mean	.63	.67	.71	.68	.63	.61	.69
SD	.02	.02	.02	.02	.02	.02	.02

interpreting generic statements flexibly in response to context, even when we ignore those who are recapitulating their own prior expectations.

The main impact of this analysis on our results is that it subtly alters our conclusions regarding Question 2: the trial-by-trial noise for experts and novices is no longer similar. Instead, we find very strong evidence that experts exhibit less trial-by-trial noise (BF_{10} of approximately 8,000 and 6,000). Evidently, much of the noise originally attributed to experts' responses is attributable to there being two modes of responding. In fact, if we remove situations in which expert participants are following the response strategy of reproducing their own priors, the total variation in interpretations is very similar between experts and novices. The primary difference is that experts' variation is systematic, and thus largely captured by the β_j parameters, while novices' variation does not exhibit any discernible pattern and is thus captured by the σ noise parameter.

In summary, the latent mixture model provides evidence that there are two ways in which experts respond to generic statements with which they disagree. When we account for these two expert response strategies, we still see that experts tend to interpret generic statements as applying more narrowly than novices do. However, we now see that novices and experts both vary their interpretations; experts vary their interpretations based on the context, whereas novices vary their interpretations unsystematically.

2.4 Discussion

Past research (Cimpian et al., 2010; Tessler & Goodman, 2019b) suggests that people who are novices within a domain might interpret generic statements about that domain differently than those who have expertise. In the present study, my collaborators and I investigated whether there are underlying patterns in how experts and novices interpret generic statements, and whether those patterns might lead to miscommunication. To do so, we asked experts and novices to interpret generic statements made by a hypothetical knowledgeable *League of Legends* player.

Regarding our first question, we found that experts would interpret generic statements flexibly. Novices were told that the teams can vary in terms of how likely they are to excel in the [early/late] game, but only experts understand *how* compositions vary, so only experts adjust their interpretations accordingly.

Regarding our second question, we found that novices would not interpret generic statements more rigidly than experts if we account for experts having two possible approaches to interpreting generic statements they disagreed with. Results from a latent mixture model suggest that experts' interpretations are less noisy than novices', but that experts' interpretations instead varied systematically. It makes sense that considering two ways in which experts might respond can greatly reduce previously unexplained noise. More noteworthy is that the added model complexity involved in accounting for those two types of behavior is supported by the data. Once we expand the model to account for plausible sources variation, we see that that experts display orderly variation based on the context, whereas novices display disorderly variation.

Regarding our third question, we found that novices interpret generic statements as more broadly-applicable than experts do. When the situation warranted it, experts' interpretations were indistinguishable from those of novices. When their interpretations deviated, it was because experts interpreted generic statements more narrowly.

On the face of it, these findings suggest that there may be systematic miscommunication when experts use generic statements to inform novices. We have found situations in which novices and experts would consistently interpret the same generic statement differently. However, there remains the question of how expert speakers decide when to use generic statements. Perhaps expert speakers are only willing to say a generic statement when they believe it applies as broadly as their audience will interpret it as applying. I begin investigating this question in the next chapter.

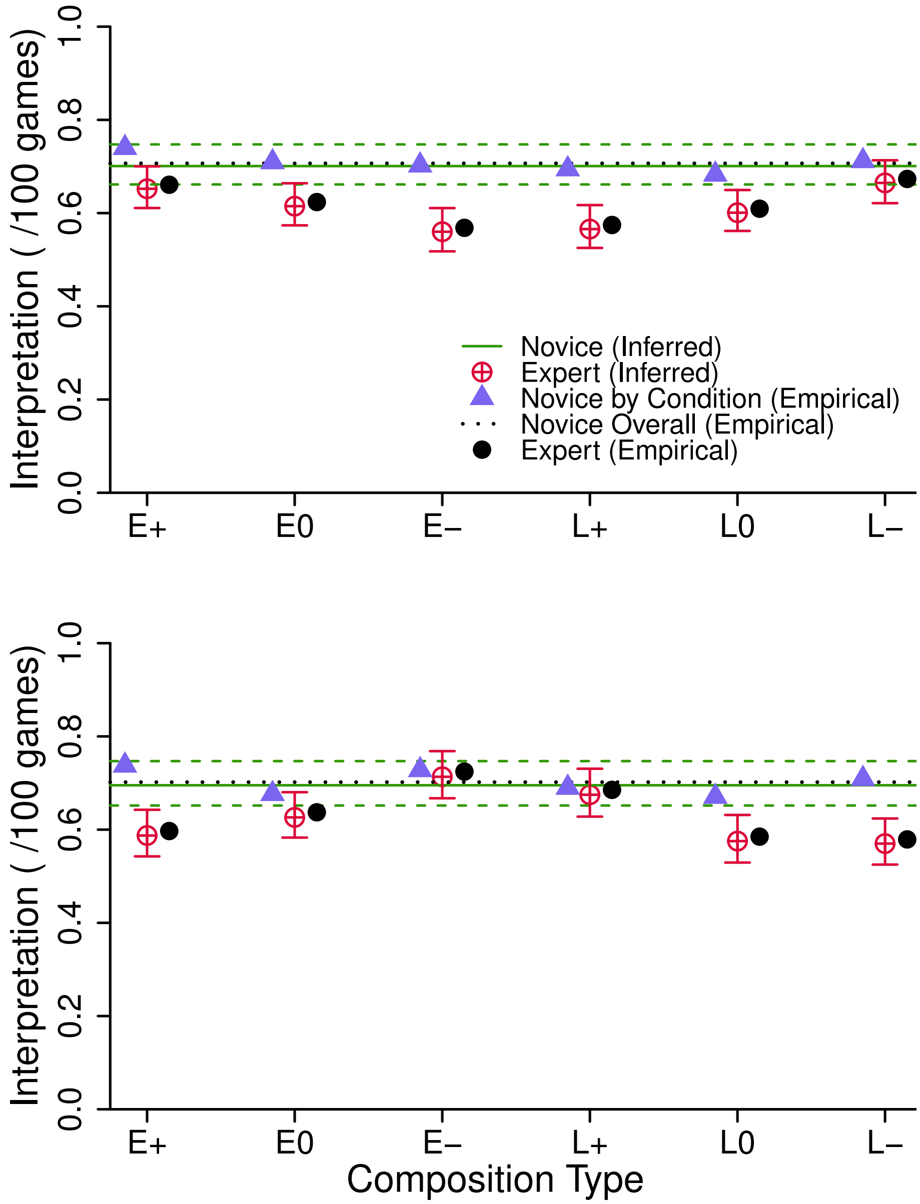

Figure 2.3: β parameter estimates for early game (top) and late game (bottom) generic statements. Error bars represent 95% credible intervals. Note that the result for β_0 is shown as a stretched horizontal line for easier comparisons.

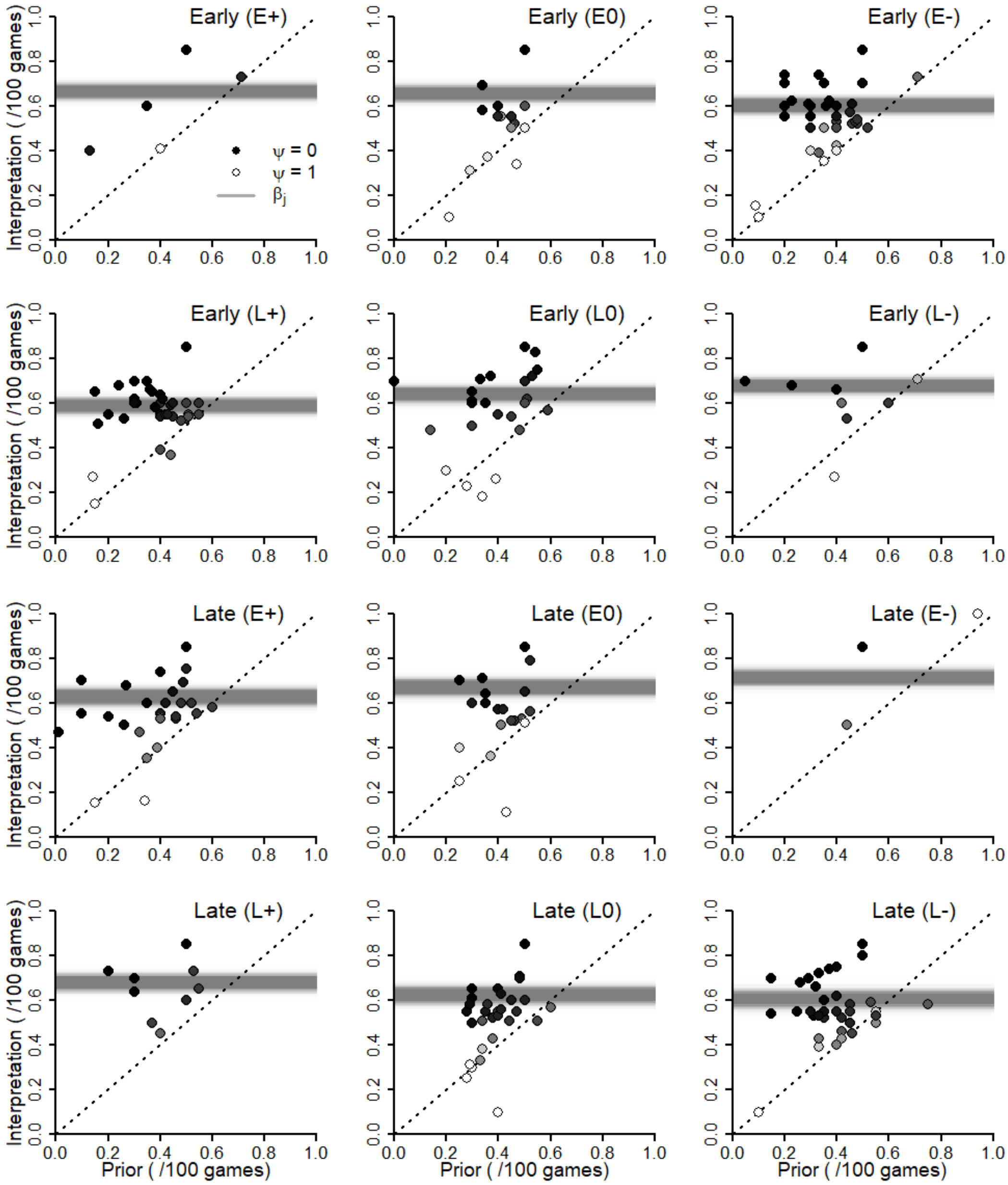

Figure 2.4: Latent mixture model inferences, accounting for two types of disagreement by experts. Each point represents an example of an expert disagreeing with a generic statement. The location of each point represents the experts' prior expectations about—and subsequent interpretation of—that generic statement. Whiter dots are inferred to be examples of an expert interpreting a generic statement they disagree with by recapitulating their own prior expectations, as indicated by proximity to the dashed $x = y$ line. The alternative type of disagreement (experts assume the disagreement is caused by the speaker seeing something unexpected) is represented by the horizontal bands, which are model inferences about β_j. The top 6 graphs show experts disagreeing with the the generic statement that each team excels in the early game, while the bottom 6 show disagreements with the generic statement that each team excels in the late game.

Chapter 3

Comparing Listeners and Speakers: Is There Miscommunication?

Abstract

The preceding chapter identified situations in which experts and novices tend to interpret the same generic statement differently. Such differences in interpretation could lead to miscommunication. However, we need to consider how expert speakers actually use generic statements. Perhaps expert speakers only use generic statements when they believe that the generic statement applies as broadly as a novice speaker would interpret it as applying. In this chapter, I describe a study in which my collaborators and I directly compare experts' beliefs in making a generic statement to how expert and novice listeners interpret those same generic statements. We find that expert-novice communication is more susceptible to miscommunication than expert-expert miscommunication. Specifically, novices tend to interpret generic statements more broadly than the speaker who made that generic statement. Future research may clarify the practical impact of such miscommunication by examining

how generic statements are used in relation to speakers' and listeners' goals.

3.1 Introduction

Generic statements are remarkably efficient because they can compress within-category variation into a simple statements of fact about that category. However, this efficiency presents a challenge because listeners must decompress them to infer the variation implied by a generic statement. As discussed in the previous chapter, listeners evidently rely on their prior knowledge to do so. Since experts and novices have different prior knowledge, there are situations in which expert and novice listeners will tend to interpret the same generic statement differently. In other words, we have situations where two people would see the same statement and differ in what message they think the statement conveys. Such situations could lead to miscommunication.

However, we also need to consider what beliefs would lead a speaker to make a generic statement in the first place. Chapter 2 shows that an expert and novice can look at the same generic statement and infer different messages, with the expert tending to interpret the generic statement as applying more narrowly. But perhaps that expert would not choose to make that generic statement themselves, believing that it applied too narrowly and would thus be misleading. Experts could and did disagree with each other regarding how broadly each presented generic statement applies, and not all experts were willing to make every generic statement they considered. It may be that experts would only have been willing to make a generic statement if they believed it applied as broadly as novice listeners did. Novice listeners would then be accurately interpreting generic statements as applying broadly, with expert listeners being overly conservative in some instances.

In this chapter, I discuss a reanalysis of the data described in the previous chapter. In it, my collaborators (Alexander Etz, Gregory Scontras, and Barbara Sarnecka) and I compare expert speakers' beliefs regarding how broadly a generic statement applies to novice and expert listeners' interpretations of the same generic statements. In doing so, we operationally define miscommunication as when a speaker and listener have a different understanding of how broadly the speaker's generic statement should be applied. Such information is often an important part of a generic statement's message. However, there are alternative conceptions of what generic statements are meant to communicate—and thus when they should be viewed as failing to do so—which are considered in the Discussion.

3.2 Methods

3.2.1 Participants

This chapter reuses the data from the preceding chapter, so we again have 49 expert participants, recruited from from *League of Legends* online forums, and 33 novice participants, recruited from UC Irvine's undergraduate participant pool. My collaborators and I only considered expert participants as potential speakers because it makes little sense to ask a novice participant whether they would make a generic statement about a game with which they have no experience. We considered both expert and novice participants as potential listeners.

3.2.2 Materials

The materials in this study are the same as those used in Chapter 1. To summarize, participants were first shown a team composition (see Figure 2.1 for an example), then asked whether they would say that the composition "excels in the [early/late] game," how often they would expect this generic statement to apply if the composition played 100 games against random opponents, and how they would interpret a knowledgeable speaker's decision about whether to make the generic statement. These team compositions were the same as the compositions used in Chapter 1. We selected them based on how well they tend to do in the early or late game. As a consequence, we see considerable variation in how likely expert participants are to make a generic statement about each composition excelling in the early or late game (see Figure 2.2).

Participants completed a total of 12 trials, each with a different team composition. For novice participants, we are only interested in how they would interpret generic statements that the knowledgeable speaker had chosen to make because we are not considering novices as potential speakers—they would presumably not have much to say about a strategical consideration in a game they have never encountered. We are also interested in how expert participants interpret the generic statements, but we are also interested in expert participants' other two responses. What prior expectations are experts expressing when they decide to make a generic statement, and how do those expectations compare to listeners' interpretations?

3.2.3 Analysis Approach

To examine how speakers' beliefs relate to listeners' interpretations, we first examined each generic statement that an expert said they would make (n = 679), noting the speaker's belief about how broadly the generic statement would apply. We think of these generic statements as ones that listeners might reasonably encounter. We then compared the beliefs behind the generic statements to 100 interpretations of those same generic statements from novices, randomly sampled with replacement from the 33 novice interpretations we collected for each generic statement in Chapter 1 (n = 396 interpretations across all generic statements). We then used the same approach to compare expert speakers' beliefs to the interpretations of expert listeners (n = 588), removing the participant who was the speaker for a given generic statement from consideration as a listener for that generic statement.

3.3 Results

The primary challenge in conducting this study is how to instantiate our operational definition of miscommunication. We have defined miscommunication as a difference between how broadly a listener and speaker believe a generic statement applies, but there is no clear threshold at which a difference becomes objectively large enough to be considered miscommunication. There is also no clear threshold at which differences are frequent enough to be a cause for concern. We instead focus on the relative frequency and magnitude of such differences when considering expert speakers are communicating with expert listeners vs. when expert speakers are communicating with novice listeners.

Examining differences between expert speakers and expert listeners on a statement-by-statement basis revealed that expert listeners do not tend to underestimate or overestimate the applicability of any particular generic statement. The differences were all centered around 0. In contrast, novice listeners tended to overestimate some items but not others, as would be predicted by the results illustrated in Figure 2.3. For some items, experts and novices are evidently aligned in how broadly they think the generic statement would apply. When they differ, it is because experts realize that the generic statement would be applied more narrowly. With this pattern in mind, the extent to which there is miscommunication becomes a question of how frequently these various types of generic statements are used. If expert speakers hardly ever make the generic statements that novice listeners would overestimate, miscommunication would not be much of an issue.

In Figure 3.1 we provide a histogram of all the differences between expert speakers' beliefs

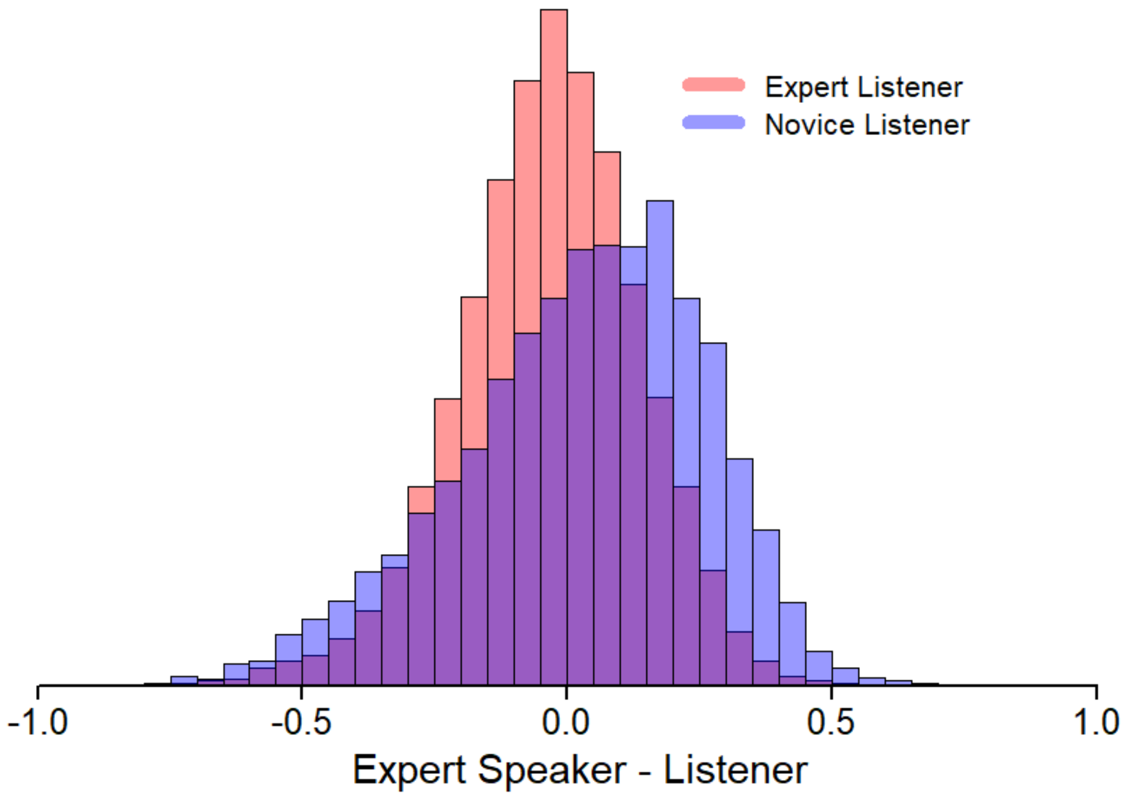

Figure 3.1: Differences between (1) expert speakers' beliefs about how broadly a generic statement applies and (2) expert listeners' (blue) and novice listeners' (red) interpretation of that generic statement.

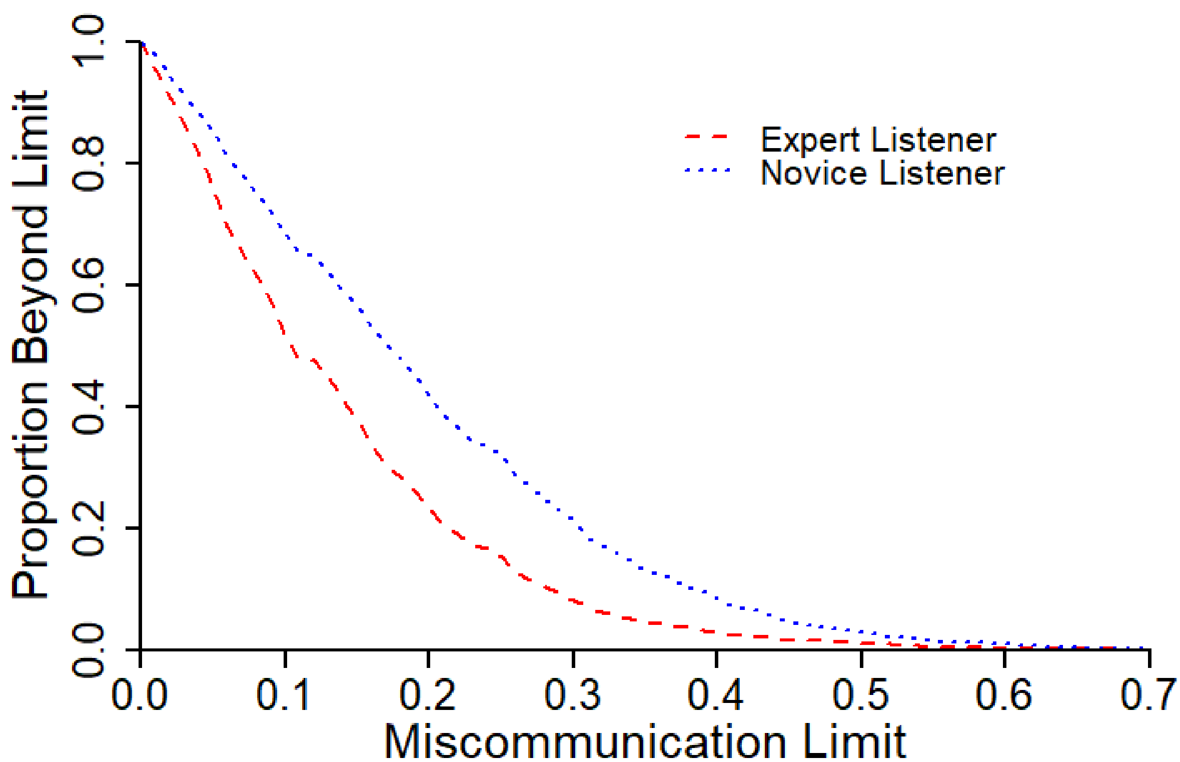

Figure 3.2: Proportion of differences between speakers and listeners (y-axis) that would be large enough to be considered miscommunication under any given threshold (x-axis).

as to how broadly a generic statement applies and expert (red bars) and novice (blue bars) listeners' interpretations as to how broadly the same generic statement applies. We include every generic statement in this histogram. Since we cycle through every generic statement that any expert participant made—and calculate the same number of differences for each generic statement—this histogram naturally incorporates the frequency with which the various types of generic statements are made. A generic statement that was made once would only contribute 100 differences to the histogram. Positive values indicate that listeners are overestimating how broadly a generic statement applies. Novice listeners' median difference from expert speakers' beliefs was 0.06, indicating that across all generic statements, they have a slight tendency to overestimate the applicability of generic statements. In comparison, expert listeners' median difference from expert speakers' beliefs was -0.01, indicating no particular tendency to overestimate or underestimate the applicability of the generic statements.

To compare the magnitude of the differences resulting from expert-expert and expert-novice communication, we considered every possible criterion for what might be a great enough difference to be considered miscommunication. Figure 3.2 provides the proportion of listener-speaker differences which would be considered miscommunication according to that criterion. For example, if our criterion is that speakers and listeners need to differ in their beliefs as to how broadly a generic statement applies by more than 1% to be considered miscommunication, then virtually every listener-speaker difference would be an example of miscommunication. If our criterion was that speakers and listeners need to differ in their beliefs by more than 60% to be considered miscommunication, then virtually none of the listener-speaker differences would be examples of miscommunication. In between those extremes, we see that a greater proportion of listener-speaker differences would be considered examples of miscommunication when the listener is inexperienced.

3.4 Discussion

In this study, my collaborators and I compared expert speakers' beliefs in making a generic statement to expert and novice listeners' interpretations of that same generic statement. We found that expert-novice interactions are more prone to miscommunication than expert-expert interactions. Specifically, novice listeners tend to interpret generic statements as applying more broadly than the expert speaker who made the statement. In contrast, expert listeners do not exhibit this tendency.

In reaching this conclusion, we are defining miscommunication as the speaker and listener failing to align their understanding of how broadly applicable a given generic statement is. The applicability of a generic statement is often an important part of its message. For example, underestimating the applicability of "mosquitoes carry malaria" could lead people to ignore a relevant danger. On the other hand, overestimating its applicability could cause a needlessly exaggerated fear of mosquitoes.

It makes sense that novices would tend to overestimate how broadly generic statements apply when we consider Figure 2.3 and Figure 2.2 in combination. Figure 2.3 shows where there are likely to be differences in how broadly experts and novices think generic statements apply. For example, there are clear differences in how experts and novices interpret the generic statement "this team excels in the early game" when it refers to an E- or L+ team composition, moderate differences when it refers to E0 or L0 team compositions, and only slight differences when it refers to E+ and L- team compositions. If expert speakers only ever say that the E+ team composition excels in the early game, there would be minimal miscommunication. Instead, Figure 2.2 shows us that, while expert speakers will make the generic statement most frequently in reference to the E+ and L- team composition, they will still occasionally make the generic statement in reference to the other team compositions. It is presumably these less commonly used generic statements that drive the miscommunication between expert speakers and novice listeners. We see the same dynamics when we look at the “this team excels in the late game” generic statements.

It is perhaps more surprising that expert listeners and expert speakers are well-aligned in Figure 3.1, particularly when we consider the latent mixture model discussed in Chapter 1. In that model, we identified situations in which expert listeners reproduce their own prior expectations when interpreting a generic statement they would not themselves make. In the context of Chapter 1, we were concerned that such behavior might suggest that an expert listener is ignoring the speaker rather than interpreting the generic statement in good faith. We found that such instances were a small but potentially noteworthy minority that did not qualitatively alter the differences we observed between expert and novice listeners' interpretations. In the present context, we might expect such interpretations to be overly conservative. As we will see in the next chapter, experts who would make a generic statement tend to think it applies more broadly than those who would not. If experts who would *not* make a generic statement assume that their own prior expectations reflect the beliefs of an expert who *would* make the generic statement, they are likely underestimating the speakers' beliefs in making the generic statement. Yet overall, we do not see evidence of expert listeners underestimating expert speakers' beliefs. Perhaps these instances are rare enough to be largely washed out in the aggregate, producing only a slightly negative median of -0.01. Alternatively, it may be that these instances are balanced out by instances of expert listeners overestimating how broadly a speaker think a generic statement applies. In any case, these are mere speculations; I would need to conduct a further experiment with a more repetitive design to draw firm conclusions about participant-level behavior.

The applicability of a generic statement is particularly important if that generic statement informs multifaceted decisions. For example, someone who is deciding whether they need to bother putting up a mosquito net must weigh the risk of disease (i.e., how broadly applicable is “mosquitoes carry malaria') against the inconvenience of a mosquito net. In this particular experiment, the generic statements communicate information about an important strategic consideration, but this consideration is only one of many. A player may need to decide between assisting one of two teammates, both of whom have characters who could be expected to "excel in the early game" To weigh a generic statement's importance relative to other considerations, listeners need a sense of how broadly it applies. We have identified situations in which that aspect of the message would be consistently distorted, leading to what could be termed a miscommunication.

Yet there are other ways of defining miscommunication, stemming from different views about what generic statements are meant to communicate. In particular, a speaker may use a generic statement to alter the listener's behavior regarding a category. For example, if the speaker's goal in saying that "mosquitoes carry malaria" is to make the listener avoid mosquitoes, then having the listener overestimate the applicability of the generic statement would be more effective than having the listener understand the generic statement in the nuanced way that the speaker does.

It may even be adaptive for inexperienced listeners to overestimate the applicability of generic statements from experienced speakers. If an experienced speaker decides that a generic statement is worth making, an inexperienced listener can safely assume that the information it contains is worth factoring into their decision-making process. From an error management perspective, erring on the side of over-applying the experienced speaker's advice is likely less costly than erring on the side of discounting it. Also, if the generic statement speaks to the only consideration for a decision (e.g., given the risk of malaria, do I avoid this mosquito or not?), overestimating the applicability may help a listener avoid sub-optimal strategies such as probability matching.

This line of thinking suggests that, when experts are using generic statements to teach novices, there may be some uncertainty regarding the question under discussion. If a generic statement is descriptive, it implicitly addresses the question of what the world is like. If a generic statement is prescriptive, it implicitly addresses the question of what the listener should do. Yet these two considerations are also fundamentally linked, since optimal behavior depends on what the world is like. Kao, Wu, Bergen, and Goodman (2014) propose a useful model for further exploring this idea, under which a listener arrives at their interpretation by simultaneously reasoning about what question a statement might be addressing and what answer it provides in the context of that implied question. I discuss this model further in Chapter 5.

To summarize, this study demonstrates that generic statements can lead to a mismatch in how experts and novices conceptualize associated statistical information. Future research can examine how such a mismatch relates to speakers' and listeners' goals.

Chapter 4

Speakers Do Not Adapt Their Use of Generic Statements to Fit Their Audience

Abstract

The previous chapter identified situations in which novices will tend to interpret a generic statement more broadly than the expert who made it. Speakers could mitigate such miscommunication by avoiding generic statements that novice listeners are likely to misinterpret. However, experts speakers may struggle to understand the perspective of a novice listener. This chapter examines whether experts adjust their use of generic statements based on the expertise of their intended audience. Results show that any such adjustments are minimal. Future research can help clarify whether expert speakers fail to appreciate that novices may interpret a generic statement differently than they do, or whether they are aware of the difference and choose to make the generic statement anyways.

4.1 Introduction

Generic statements are particularly valuable as a way of introducing people to topics with which they are unfamiliar. They reference a category rather than specific individuals, licensing listeners to make inferences about previously unencountered individuals (Gelman et al., 2002). Adults seem to grasp the pedagogical value of generic statement, at least implicitly, because they use generic statements frequently in child-directed speech (Gelman et al., 1998; Pappas & Gelman, 1998; Gelman et al., 2008). Children also seem to appreciate the pedagogical value of generic statements on some level, since they actively seek out generic statements (Cimpian & Park, 2014) and privilege generic statements in memory (Cimpian & Erickson, 2012).

The previous chapter identified situations in which generic statements would lead to miscommunication in a pedagogical context. Specifically, novice listeners tend to interpret generic statements as applying more broadly than do the expert speaker who made them. However, expert speakers were not told about the expertise of their audience. Instead, they were just asked if they would make a generic statement when explaining the referenced team composition "to a friend." Perhaps expert speakers would behave differently if they knew that they were speaking to a novice. If expert speakers know they are speaking to an novice audience, they might only use broadly applicable generic statements, dismissing more narrowly applicable generic statements as too nuanced for novice listeners to worry about. And if expert speakers do make such an adjustment, novice listeners would be fully justified in interpreting generic statements as broadly applicable.

Yet we might also expect expert speakers to struggle with adapting their behavior to account for the fact that they know things which their audience does not. Fundamentally, it can be challenging for people to put themselves in the shoes of someone with less experience. This difficulty is referred to as the "curse of knowledge" (Camerer, Loewenstein, & Weber, 1989). The original study which described it in such terms focused on economics, finding that informing only some participants about the true value of an item led to those participants to make pricing decisions as if every participant had access to that privileged information. Since then, researchers have examined how the curse of knowledge plays out in communication. They have consistently found that participants who have privileged information that disambiguates the meaning of an statement overestimate how clearly a participant without such information will be able to interpret the statement. Even professional educators struggle to remember what it was like to not know something, tending to teach in a way that fits with their current understanding of the material rather than that of their students. With this phenomenon in mind, we might expect expert speakers to neglect a listener's lack of experience when evaluating whether a generic statement is worth making.

A closely related line of research investigates how and when speakers design utterances to fit their audience (Clark & Murphy, 1982; Bell, 1984). As with the "curse of knowledge," research examining "audience design" suggests that speakers will struggle to adjust their use of generic statements based on their audience's expertise (or lack thereof). Tailoring language to fit a specific audience involves realizing that the audience has a different perspective, applying a theory of mind to identify the relevant features of the audience's perspective, and determining what sort of communication would be effective in that context (Sulik & Lupyan, 2018). Children struggle to demonstrate this capacity, long past the point at which they can demonstrate that they have developed a theory of mind (Arvidsson, Pagmar, & Uddén, 2022). Even in simplified situations, speakers' ability and/or motivation to make adjustments seems to depend on them getting feedback in the moment that blatantly demonstrates they are not being understood (Gann & Barr, 2014; Jucks, Becker, & Bromme, 2008; Yoon & Brown-Schmidt, 2019). The experiment on which this chapter is based does not allow for any direct interaction between speakers and listeners, so speakers would have to preemptively realize that their generic statements might be misinterpreted.

This chapter focuses on two key questions: (1) do experienced speakers adjust their use of generic statements in response to the expertise of their audience and (2) if so, is it sufficient to preclude miscommunication? To investigate these questions, my collaborators (Alexander Etz, Gregory Scontras, and Barbara Sarnecka) and I conducted an experiment in which some expert participants were asked whether they would make a generic statement when speaking with a fellow expert, while other expert participants were asked the same question in relation to speaking to novices. We also developed a descriptive modelling approach that allows us to quantify how participants' behavior changes based on the expertise of their audience. In performing this study, we again operationally define miscommunication as a speaker and listener failing to align their understanding of how broadly a generic statement applies, and we discuss the implications of alternative definitions. Speakers might prevent such miscommunication by avoiding generic statements that their audience would interpret differently. However, speakers might not realize that their audience will interpret some generic statements differently than they would, as predicted by the curse of knowledge.

4.2 Methods

4.2.1 Participants

We recruited expert participants (n=75) from online *League of Legends* forums. To ensure that expert participants were well acquainted with the domain, we only included those ranked in the 27th percentile (Silver tier) of players or higher. This sample was mostly men (84.1%).

We also recruited novice participants (n=25) from UC Irvine's undergraduate participant pool. We excluded 10 additional participants because they indicated some experience with *League of Legends*. This sample is mostly women (80%).

Figure 4.1: Matchup example. Note that, in this condition, the participant's intended audience is a new player. "Laning phase" refers to the early part of a game.

Unlike the experiment described in Chapters 2 and 3, the materials for this experiment asked participants to reason about matchups between two opposing characters (Figure 4.1) rather than an entire team composition. More specifically, we asked participants to think about which of the two characters would likely have an advantage in the early phase of a game. Such statements are generic statements because some characters have inherent advantages over others, but there are myriad other variables that can impact how a given game will play out. We switched from discussing entire compositions to discussing individual matchups because expert participants indicated that reasoning about a composition took considerable time and effort. This alteration also allows us to see if the findings from Chapters 2 and 3 conceptually replicate to a slightly different generic statement.

As in that previous experiment, participants were first given a basic overview of *League of Legends*, including definitions of key terms related to the game (e.g., phases of the game, resources such as gold). Expert participants were then presented with a series of matchups. For each matchup, they were asked about which of the two characters gets more gold in the laning (i.e., early) phase of the game. Because collecting more resources confers an advantage, and gold is a fundamental resource, collecting more gold was our operational definition of gaining an early advantage.

First, expert participants were asked whether they would endorse a generic statement saying that a given character "gets more gold" in the early game. They could answer "yes," "no," or "I don't know." We coded "I don't know" responses as participants not making the generic statement. Between subjects, we varied the expertise of the audience to which the generic statements would be directed. One group of participants was asked whether they would make the generic statements in explaining the matchup to a "new player." The other group was asked about generic statements made to an "experienced player." To emphasize this manipulation, the audience's experience was highlighted and italicized.

Next, expert participants were asked to estimate how often they would expect the referenced character to get more gold in the early game if the matchup was played 100 times. Finally, expert participants were told that a knowledgeable speaker had made the generic statement based on having seen the matchup play out 100 times. Expert participants were asked to interpret the generic statement in terms of how often they thought the speaker had seen the referenced character get more gold in the early game.

Novice participants were only asked to interpret generic statements. All participants were asked to describe their playing experience. Participants who indicated that they had prior experience were asked to provide more detail in terms of their ELO ranking (Elo, 2008), time played, knowledge of the characters included in the experiment, and how frequently they play in the types of matchups under discussion.

We removed expert participants who said they would make a generic statement that they think applies to fewer than 10% of examples, and those who said that they would not make a generic statement that they think applies to more than 90% of examples. We did so to

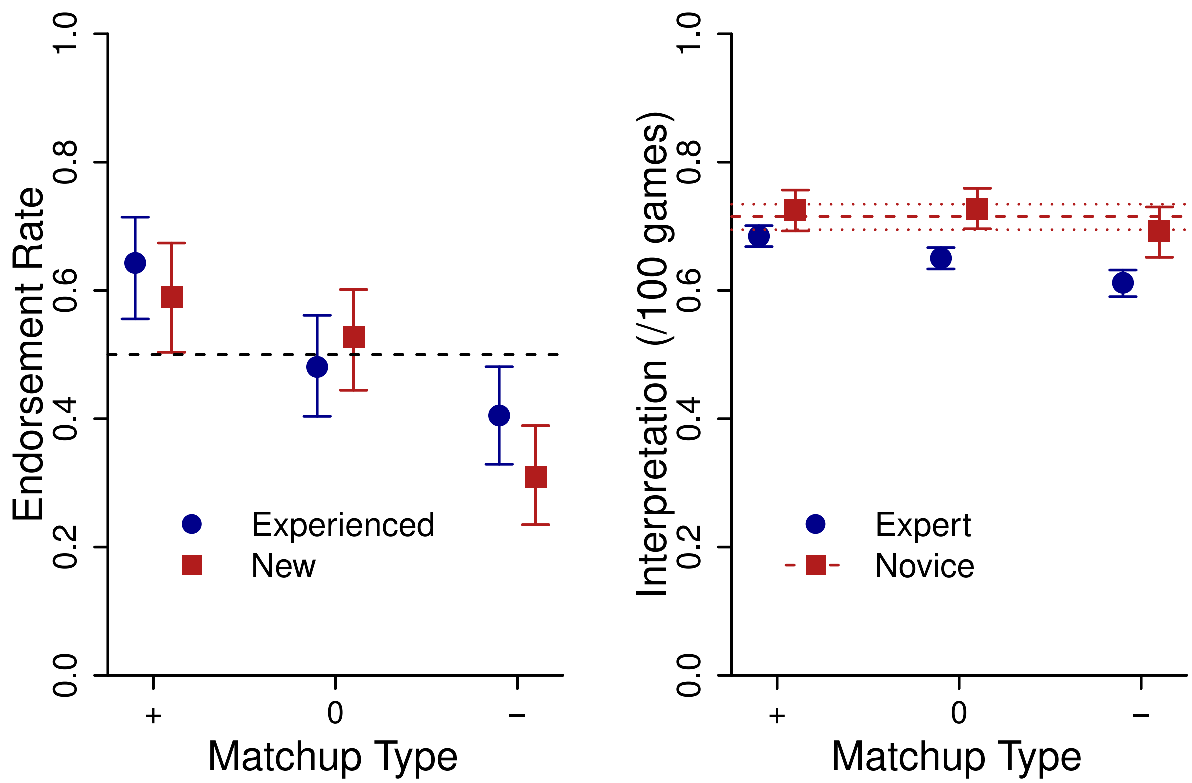

Figure 4.2: **Left:** Proportion of expert participants who would endorse the generic statement that the referenced character would excel in the given matchup regardless of intended audience. Symbols +, 0, and − designate referenced character's advantage (or lack thereof) in the given matchup. **Right:** Expert and novice participants' interpretations of generic statements by matchup type. All error bars are 95% bootstrapped confidence intervals.

prevent such outliers from having an outsized impact on our model. These criteria excluded 7 additional expert participants.

Design

Participants viewed 12 matchups in total. In some of the matchups, one character had a distinct advantage, while others were more evenly matched. In even matchups, it made no difference which character was referenced in the generic statements, so participants only saw 4 such matchups. For the imbalanced matchups, the generic statements referenced the character with a clear advantage on half of trials, and referenced the character with a clear

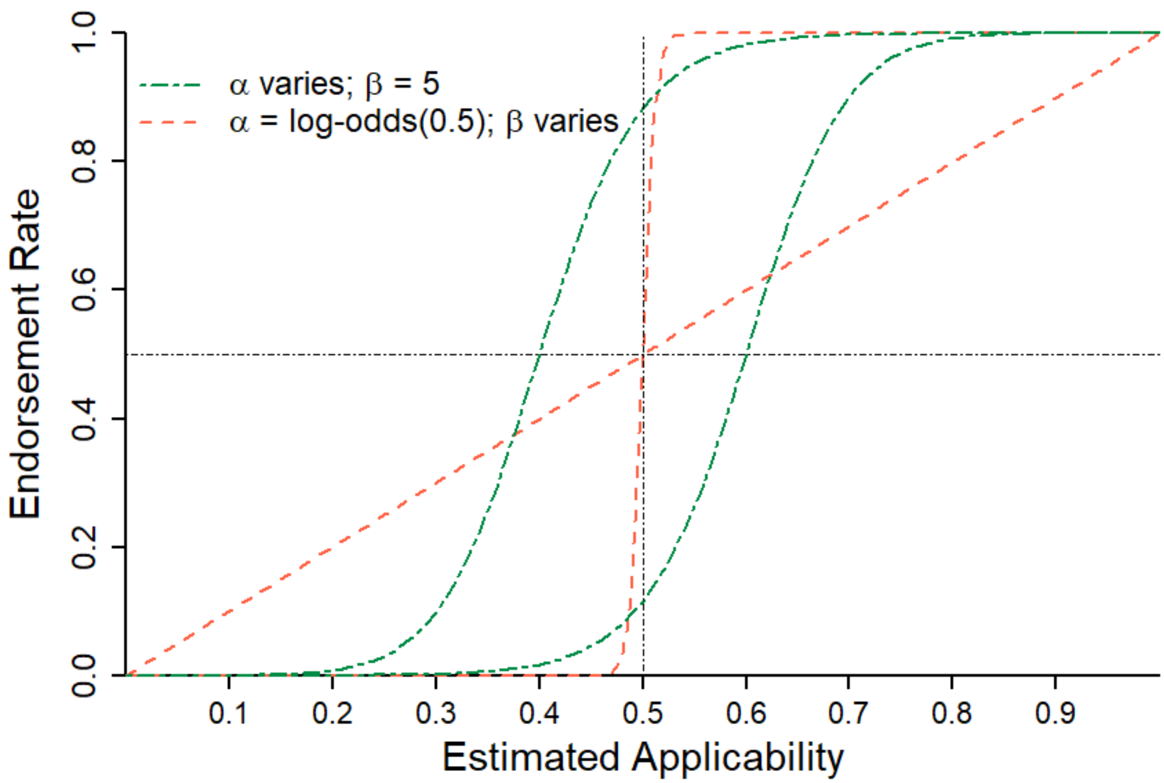

Figure 4.3: Illustrative example of logistic model varying in terms of shift, α = log-odds$(0.4, 0.6)$, and scale, $\beta = (1, 45)$.

disadvantage on the other half. Participants saw 8 such matchups.

The left panel of Figure 4.2 demonstrates that experts understand the matchups as we had intended. As the matchup under discussion becomes more disadvantageous for the referenced character, experts become less likely to make a generic statement about that character excelling in the matchup.

This experiment can be summarized as a 3 (character advantage: +, 0, or $-$) x 2 (listener type: experienced or new player) design. Character advantage was varied within subjects; listener type was varied between subjects. The character advantage conditions are not our unit of analysis because experts might reasonably disagree with each other as to the extent to which a specific character has an advantage. Our aim is to descriptively model the process by which expert speakers decide when to endorse a generic statement given their prior expectations. We manipulate character advantage to ensure that the prior expectations (i.e., the model inputs) experts consider when deciding whether to make a generic statement are sufficiently varied for us to capture the bounds of that process.

4.3 Results

4.3.1 Comparing Interpretations

We first replicated the findings of Chapter 2, which showed that novices tend to interpret generic statements as broadly applicable regardless of the context, whereas experts can use their prior knowledge to distinguish between broadly- and narrowly-applicable generic statements, as shown in the right frame of Figure 4.2. Adopting the same model comparison methodology, we found that the best performing model assumed that experts varied their interpretations by condition and that novices did not. We also again found that experts and novices have similar trial-by-trial noise until we consider the latent mixture model, which reduces the noise ascribed to experts' interpretations. These results suggest that there could indeed be miscommunication because there are generic statements that experts and novices would interpret differently, specifically those that experts consider to be more narrowly applicable. However, there is still the possibility that expert speakers would not use narrowly-applicable generic statements when speaking to novices, avoiding the generic statements in which their own interpretations would differ markedly from that of their audience.

4.3.2 Modeling Approach

To examine whether expert speakers alter their process for deciding when to use generic statements, we first need a working model of by which experts decide to make generic statements. Unlike past efforts to model this process (e.g., Tessler & Goodman, 2019a), we want a model that is more descriptive than cognitive because we are looking to summarize people's behavior such that we can look for informative patterns. To our knowledge, the only previous effort to descriptively quantify the process by which people decide to make a generic statement did so by taking the mean of participants' prior estimates for generic statements they endorsed (Cimpian et al., 2010). While this approach was adequate in the context of the specific purpose for which it was used, it neglects information provided by generic statements that a participant rejected.

We instead used a hierarchical model in which a speaker's decision to make a generic statement is a binary choice, made with some probability θ. This probability is in turn determined by a function which takes as its input the speaker's estimate of how broadly applicable the generic statement is. Our primary interest is in characterizing how this functional relationship compares to listeners' interpretations.

In selecting a specific functional form, we want to capture the important aspects of the relationship without making unnecessary assumptions. First, we assume that the probability of a generic statement being endorsed will increase monotonically from 0 to 1 as a function of the speaker's belief of how broadly applicable it is. We also assume that the function modeling this process can vary along two dimensions: shift and scale. The shift parameter determines the threshold at which a speaker's beliefs make them more likely than not to endorse the generic statement. The scale parameter determines the steepness of the function around the threshold. In other words, it controls how strict that threshold is.

These assumptions together suggest a two-parameter sigmoidal function. We chose a logistic model, which can be specified as

$$\theta = \frac{1}{1 + e^{-\beta(x-\alpha)}}, \tag{4.1}$$

where θ is the probability of endorsing a generic statement, and x is the log-odds of the speaker's belief as to how broadly applicable the generic statement is. With this parameterization, α controls the location of the threshold (i.e., shift) and β controls the steepness of the curve (i.e., scale). As demonstrated in Figure 4.3, the β scale parameter can create a step function, corresponding to a strict threshold. At the other extreme, it can create a consistent and gradual slope.

We allow the functional relationship to differ depending on the audience, with separate group-level parameters α_{new} and β_{new} for the new-player (inexperienced) audience versus α_{exp} and β_{exp} for the experienced audience. We test this assumption in investigating the first question below, by comparing the group-level parameters of each condition.

To account for individual differences, we also allow the functional relationship to vary across participants. In the model, the ith participant's function is specified by offsets from the group-level parameters described above. We denote these offsets as α_i and β_i. We assume that each α_i is independently drawn from either $N(0, \sigma^2_{\alpha exp})$ or $N(0, \sigma^2_{\alpha new})$ depending on the audience condition to which the participant was assigned. We assume that each β_i is independently drawn from $N(0, \sigma^2_{\beta exp})$ or $N(0, \sigma^2_{\beta new})$, again depending on audience condition. We split these σ parameters by condition to account for the possibility that one audience condition leads to more participant-level variation than the other. If participant i was assigned to the new-player audience condition, their functional relationship would be

$$\theta_i = \frac{1}{1 + e^{-(\beta_{new}+\beta_i)(x-\alpha_{new}-\alpha_i)}}. \tag{4.2}$$

Table 4.1: Model inferences for key parameters. α_{exp} and α_{new} have been transformed using $\frac{1}{1+e^{-\alpha}}$ so that we report the estimate of speakers' threshold on the scale of the generic statement's estimated applicability, rather than on a log-odds scale.

Parameter	Posterior Mean	95% Credible Interval
$\frac{1}{1+e^{-\alpha_{exp}}}$	0.52	(0.50, 0.54)
$\frac{1}{1+e^{-\alpha_{new}}}$	0.54	(0.53, 0.56)
β_{exp}	3.43	(2.57, 4.50)
β_{new}	6.99	(4.67, 10.21)
$\sigma_{\alpha exp}$	0.18	(0.03, 0.36)
$\sigma_{\alpha new}$	0.16	(0.05, 0.26)
$\sigma_{\beta exp}$	1.60	(0.70, 2.65)
$\sigma_{\beta new}$	3.49	(1.57, 6.26)

We implemented the model in JAGS (Plummer, 2003). We assigned both α shift parameters (α_{exp} and α_{new}) a prior distributions of $N(0, 2^2)$, which is agnostic about whether α is positive or negative. Since we are using a log-odds scale, an α value of zero corresponds to 0.5 on the response scale (see Figure 4.3). We assigned both the β scale parameters (β_{exp} and β_{new}) a prior distribution of $N^+(5, 20^2)$. Most of the mass of this prior is between 1 and 45. A β value of less than 1 alters the functional form in a way that is possible but unlikely for this context. A value of 45 is a soft upper bound for β, creating what is essentially a step function. For the parameters governing the distributions of participant-level α_is ($\sigma_{\alpha exp}$ and $\sigma_{\alpha new}$), we assign prior distributions of $N^+(0, 1^2)$. Finally, for the parameters governing the distributions of participant-level β_is ($\sigma_{\beta exp}$ and $\sigma_{\beta new}$), we assign prior distributions of $N^+(0, 10^2)$.

Figure 4.4 shows the inferred group-level functional relationships alongside summaries of the empirical data. Experts either would or would not make the generic statement, but we have again binned those binary responses based on participants' estimates as to how broadly the given generic statement would apply. Each point represents the rate at which experts would make a generic statement if they believed its applicability fell within that range (0 to 0.1, 0.1 to 0.2, etc.). For example, no participants endorsed generic statements they believed would only apply 10-20% of the time. Again, we excluded participants who

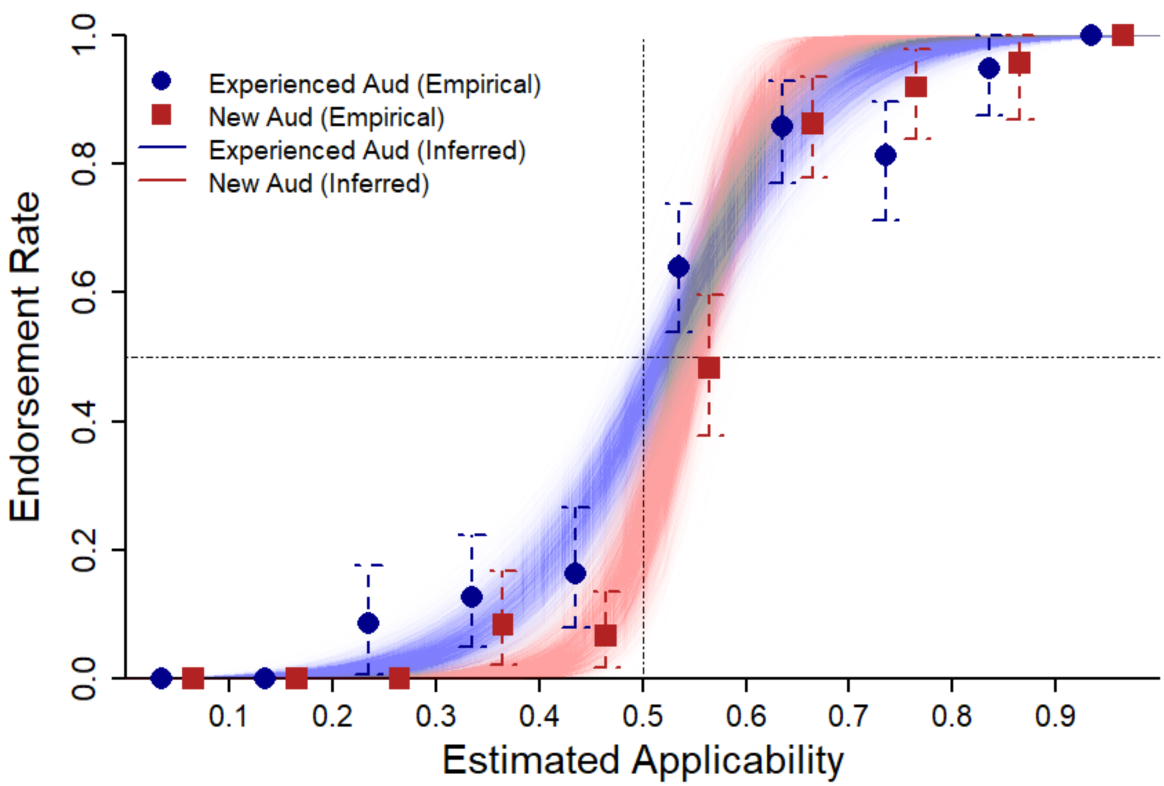

Figure 4.4: Rate at which experts endorse the generic statement that the referenced character excels in the given matchup as a function of the participant's expectations about how broadly the generic statement applies. Red and blue sigmoid lines represent a random sample from each respective condition's joint posterior distribution. Error bars are 95% bootstrapped confidence intervals.

endorsed a generic statement which they only expected to apply 0-10% of the time and those who did not endorse a generic statement which they expected to apply 90-100% of the time; we interpreted such responses as contaminant behavior.

We present group-level estimates for parameters of interest in Table 4.1. The α estimates suggests that experts become more likely than not to endorse a statement when their estimated applicability of the statement is slightly higher than 50%, whether they are speaking to a new or experienced player. The β estimates suggests the decision threshold tends to be rather steep in both conditions. The α and β parameters may trade off with each other in the model's posterior distribution, so inferences about one or the other may not fully capture the functional relationship of interest. Since these parameters jointly determine the functional relationship between estimated applicability and endorsement probability, we will look at them jointly next.

4.3.3 Question 1

Our goal in creating this model is to examine whether experts adjust their use of generic statements based on the expertise of their audience. The basic model uses two parameters, shift and scale, to characterize the use of generic statements, so experts can potentially adjust their use of generic statements along two dimensions. There are thus four possibilities for how the curves might compare between audience conditions: the shift changes, the scale changes, both change, or neither change.

Table 4.2 lists the evidence in favor of each model using Jeffreys weights – an extension of the Bayes factor for situations involving more than two models (Vandekerckhove, Matzke, & Wagenmakers, 2015).[1] If experts use similar thresholds for both audiences, the data

[1] We derived the Jeffreys weight J_i for models $i = \{1, 2, 3, 4\}$ from the series of Savage-Dickey Bayes Factors for nested comparisons against model 4 (the full model), such that $J_i = BF_{i4} / \sum_j BF_{j4}$.

Table 4.2: Model summaries and evidence.

Model #	# of αs	# of βs	Jeffreys Weight
1	1	1	0.39
2	2	1	0.01
3	1	2	0.57
4	2	2	0.03

would support more parsimonious models which only have a single α shift parameter that applies to both conditions. If experts are similarly strict in following the threshold for both audiences, the data would support models with a single β scale parameter that applies to both conditions.

Our data provide the most support for Model 3, which specifies separate scale parameters for each audience but only a single shift parameter. We also find considerable evidence in favor of Model 1, which specifies a single shift and single scale parameter. The evidence for Model 3 over Model 1 is minor, $BF_{31} = J_3/J_1 = 1.5$.[2] There appears to be a difference in scale when considering Figure 4.4, but there also remains much uncertainty in the estimates of β_{new} and β_{exp}. We find scant evidence in favor of either of the models which specify separate shift parameters. Overall, our data indicate that experts may make a slight adjustment to their use of generic statements based on their audience, but only in terms of the slope of the threshold at which they become more likely than not to make a generic statement, not the location of that threshold.

[2]More informed priors on the β parameters, such as $N^+(5,5)$, would lead to somewhat stronger evidence for Model 3. However, our choice of diffuse priors reflected our limited knowledge of this novel experimental context, and using more informative priors would not have changed our qualitative inferences.

4.3.4 Question 2

Our second question is whether any adjustments participants did make based on the expertise of their audience would be sufficient to prevent miscommunication. As in Chapter 3, we operationally define miscommunication as a speaker and listener failing to align their understanding of how broadly a generic statement applies.

In terms of our model, reducing such miscommunication would manifest primarily as $\alpha_{new} > \alpha_{exp}$. In other words, speakers are less willing to use narrowly-applicable generic statements when speaking to novices, since novices would tend to misinterpret those generic statements. It could also be helpful for the slope to get steeper ($\beta_{new} > \beta_{exp}$) as such an adjustment indicates that speakers are more rigid in their determinations of which generic statements are and are not worth making. However, α is the main parameter of interest because adjusting the scale would only help reduce miscommunication if the shift were also adjusted.

Based on our analysis for Question 1, we have strong evidence that the value of α does not change in response to the audience condition. If, for the sake of argument, we assume that there is a difference in the α values and simply test how likely that difference is to be greater for the new-player audience condition, we get a Bayes Factor of approximately 10 in favor of $\alpha_{new} > \alpha_{exp}$. Such a shift is indeed in the direction that would help reduce miscommunication, but the shift is evidently slight, if it exists at all.

Our analysis for Question 1 indicated minor evidence that the value of β changes in response to audience condition. Another directional hypothesis test, this time for β, produces a Bayes Factor of approximately 183 in favor of $\beta_{new} > \beta_{exp}$, which would indicate that experienced speakers are more strict in deciding whether a generic statement is worth making if they are speaking to an inexperienced audience. However, since there is hardly any change in α, such an adjustment to β would do little to close the gap between the expectations of the speaker and the listener.

4.4 Discussion

In this study, my collaborators and I examined whether speakers adjust their use of generic statements based on the expertise of their audience and whether such adjustments are sufficient to avoid miscommunication. We found evidence that experts do not adjust their use of generic statements, leaving room for miscommunication as we have operationally defined it.

Interpreting null results is difficult. Perhaps participants did not demonstrate impactful adjustments because our manipulation was ineffective, even though we emphasized the experience (or lack thereof) of the intended audience. In particular, our participants did not have a conversational partner to actively provide cues as to how the generic statements were being interpreted. Past research on how speakers design utterances to fit the expertise of their audience (e.g., Sulik & Lupyan, 2018) indicates that speakers use such cues to adapt over the course of a conversation. Nevertheless, our findings indicate that speakers do not adjust their use of generic statements when they are explicitly told that the statements will be directed towards a novice. As such, simply informing experts about the expertise of their audience is insufficient for avoiding the potential miscommunication identified in Chapter 3.

Further investigation is needed to determine whether speakers fail to adjust their use of generic statements when speaking to novices because of the curse of knowledge. Past research suggests that the curse of knowledge is quite difficult for speakers to overcome even when they are warned about it (Damen, 2018). We might reasonably assume that it presents difficulties for speakers in the present experiment as well, but based on these results alone, we cannot tell whether our expert participants are cognizant of how expert and novice listeners will interpret generic statements differently. They might struggle to imagine how they would interpret a generic statement if they were a novice, but they might also be aware of the difference and decide that a generic statement is still worth making.

As discussed in Chapter 3, it may actually be adaptive for listeners to overestimate how broadly an expert's generic statement applies. It thus remains an open question as to whether expert speakers *should* adjust their use of generic statements when speaking to novices. This question is separate—although potentially related—to that of *why* speakers are evidently not adjusting their use of generic statements.

As was also discussed in Chapter 3, speakers may implicitly use generic statements to address different questions when speaking with experts and novices. When speaking to fellow experts, speakers might look to communicate the statistical nuances of a category, which the expert listener can use to fine-tune their decision making. When speaking to novices, speakers might focus more on directly prescribing what the listener should do in regards to the category. In the present experiment, the most pronounced alteration to speakers' use of generic statements is that their criteria for when to make a generic statement may become stricter (i.e., exhibit a steeper boundary between generic statements they would and would not make) when speaking to novice listeners. Such an adjustment could indicate that speakers are indeed focused on novice listeners' behavior and are therefore reducing the world into information that the novice listener should or should not act on. In doing so, speakers may ignore statistical nuances, assuming that novice listeners are not yet capable of the complex decision-making that would make such nuances valuable.

Chapter 5

Conclusions

Abstract

In this dissertation, I have pointed to situations in which experts and novices interpret the same generic statement differently. In spite of these differences, speakers do not adjust their use of generic statements, even when told that their audience is a novice. As a consequence, novices are prone to miscommunication, tending to overestimate how broadly generic statements apply. In this final chapter, I highlight some of the recurring themes from this dissertation and discuss opportunities to develop them further. I focus on what we have learned and what remains unresolved regarding the semantics, purpose, developmental trajectory, and real-life implications of generic statements. In considering future research, I return to the theme of Chapter 1: balancing ecological validity and experimental control.

5.1 Semantics of Generic Statements

Throughout this dissertation, I have remained agnostic regarding the semantics of generic statements. I only assume that an important part of the message being communicated is how broadly the generic statement applies (i.e., its implied prevalence). Based on this assumption, I have been investigating how that implied prevalence is communicated in a pedagogical context.

Yet the research outlined in this dissertation is pertinent to the debate about the underlying semantics of generic statements. As described in the introduction, there are currently two leading accounts. One account focuses on outlining the criteria under which a generic statement is true. The other account proposes that a generic statement is considered true if it would more closely align a hypothetical novice listener's understanding of the environment with that of the speaker. Both accounts suggest that listeners interpret the implied prevalence of generic statements by using their own prior knowledge to reverse engineer the speaker's thought process in deciding to make a generic statements.

These accounts are difficult to disentangle because they make overlapping predictions, but they reach those predictions from opposite directions. For example, both accounts predict that speakers will make generic statements that reference dangerous and distinctive traits, even if the trait is rare within the referenced category (e.g., "mosquitoes carry malaria"). Under the implied prevalence account, dangerous traits are assumed to be relatively unique, and distinctive traits are unique by definition. Given the rarity of such traits, a hypothetical novice would tend to underestimate their prevalence. A generic statement referencing such traits would be worth making because it would nudge their understanding closer to the truth. Under the criteria-based account, dangerous or distinctive referenced traits are explicitly included as one of the criteria by which a speaker can endorse a generic statement, regardless of the implied prevalence.

Similarly, both accounts predict that, in the absence of a causal link between a referenced trait and category, speakers will still tend to endorse a generic statement that applies to at least 50% of category members (e.g., "cars have radios"). Under the implied prevalence account, the generic statement would be useful because a hypothetical novice would have no reason to suspect that the trait is so prevalent, given the lack of a causal connection between category and trait. The criteria-based account again makes this an explicit criteria by which a speaker would endorse a generic statement. However, the criteria based account struggles because such an explicit rule makes it difficult to explain situations in which a speaker would *not* endorse such a generic statement. For example, most speakers would find a statement like "books are paperbacks" to be a bit strange, even though the majority of books are indeed paperbacks. In contrast, the implied prevalence account can accommodate such situations by proposing that the speakers would not expect the statement to meaningfully improve a hypothetical novice's expectations. In the case of this example, the implied prevalence account suggests that a speaker would conceptualize a hypothetical novice listener as already expecting most books to be paperback. The hypothetical novice would thus not be helped by the statement.

Both accounts would benefit from further research. While the implied prevalence account has a quantitative model that makes fairly specific predictions, it hinges on speakers' and listeners' prior expectations, which can be vague and difficult to measure. How do we determine what a speaker thinks a hypothetical novice would expect? Similarly, while the criteria-based account makes fairly specific predictions, the criteria it hinges on can be vague and based on subjective judgements. When is a trait sufficiently concerning to qualify as dangerous? When is a trait sufficiently unique to qualify as distinctive? Most of the experiments directly examining these accounts have focused on establishing the plausibility of each proposed mechanism (Tessler & Goodman, 2019a, 2019b; Leslie, Khemlani, & Glucksberg, 2011). The account under consideration typically acts as the hypothesis to be tested, so methodological choices incorporate implicit assumptions about which factors might be relevant to people's judgements about generic statements.

The descriptive model developed in Chapter 4 provides a method for analyzing how people decide whether generic statements are true without first having to commit to a conceptual framework. The model conceptualizes a speaker deciding whether to endorse a generic statement as a probabilistic event, where the probability of endorsing a generic statement changes as a function of the speaker's belief about the referenced trait's prevalence in the referenced category. The primary assumption of this model is that, all else being equal, speakers will be more likely to endorse a generic statement if they believe it applies more broadly. Based on this assumption, the model looks for a threshold at which speakers become markedly more likely to endorse a given generic statement, allowing us to compare speakers' decisions in terms of the location and slope of that threshold.

The model also allows us to examine individual differences in that decision-making process. This feature can help identify statements that might be controversial, with some speakers endorsing and others rejecting them. In doing so, we can examine how the vague aspects of each competing semantic account might play out in real-life communication. Such controversies could be due to differences in how people conceptualize the referenced trait, as predicted by the criteria-based account, or how they conceptualize a hypothetical novice, as predicted by the implied prevalence account.

Importantly, this descriptive model does not commit to the assumptions of the implied prevalence model, even though it does emphasize speakers' beliefs about implied prevalence. The implied prevalence account focuses on cognitive processes, modelling speakers' reasoning as being centered on their conceptions of a hypothetical listener's prior knowledge regarding the topic of a generic statement and how it would be updated in light of that generic statement. As such, experiments inspired by this account have focused on capturing these conceptions, since they are assumed to be the driving force behind the semantics of generic statements (Tessler & Goodman, 2019a, 2019b). In contrast, this descriptive model simply looks to establish what implied prevalence would be sufficient for speakers to endorse a specific generic statement.

The key methodological decision in implementing this descriptive model is deciding which generic statements should be thought of as being the same type. In Chapter 4, my collaborators and I assume that all generic statements about Character A getting more gold in the laning phase are fundamentally the same type of generic statement, regardless of the specific identity of Character A. Based on this assumption, we vary the identity of Character A to manipulate the implied prevalence of the generic statement under consideration, allowing us to establish what implied prevalence would prompt a speaker to make that type of generic statement. Since we are interested in how speakers adjust their use of generic statements in response to their audience, we assume that generic statements directed towards novices were of a fundamentally different type than those directed towards experts. Based on this assumption, we apply the model separately based on who the intended audience is. Our subsequent analysis indicates that this second assumption is questionable, because we find minimal differences between generic statements directed towards novices and experts. The point is that we can use this model as a framework for testing whether any variable—be it the content of a generic statement or the context in which it is made—makes a difference in how speakers decide when a generic statement applies broadly enough to be worth making. Given the proposals of the two leading semantic accounts, we might be particularly interested in examining how speakers' behavior is impacted by the nature of the referenced trait and category, the causal link between the trait and category, and the expectations of a hypothetical novice.

5.2 Purpose of Generic Statements

On the face of it, these two accounts of the semantics of generic statements could be viewed as pointing to a distinction without a difference. Under both accounts, listeners use their prior knowledge to make pragmatic inferences about the implied prevalence of a generic statement, so the research presented in this dissertation is largely compatible with both accounts. Furthermore, the criteria outlined by the first account are essentially a catalog of the types of situations in which the second account predicts a generic statement would be worth making.

On closer inspection, they differ in their points of emphasis. The criteria-based account views the implied prevalence of a generic statement as a secondary consideration. It places more emphasis on whether the generic statement highlights an aspect of the environment that is important, be it the danger posed by a referenced trait or the predictive power of linking a trait and category.[1] In contrast, the second account views the implied prevalence as the primary consideration, with the importance of the generic statement to the listener being more of a byproduct.

These differing points of emphasis speak to the idea that generic statements can be used both to inform a listener about the environment and to prescribe a course of action. These considerations are fundamentally intertwined. Providing information about an environment allows listeners to make inferences about the best course of action, and advocating specific actions allows listeners to make inferences about the environment. As with the semantic accounts, the difference is one of emphasis. Indeed it may be that each semantic account describes a purpose for which generic statements can be used. Is the generic statement focused more on informing the listener or on influencing their behavior?

Both the speaker and listener play a role in determining a generic statement's purpose. A speaker may have a particular goal in mind when making a generic statement, and a listener may be looking for a particular type of information when interpreting the generic statement. In aligning their conceptions of a generic statement, speakers and listeners are negotiating the

[1]Distinctive traits are predictive in that they reliably distinguish the referenced category from other categories. Traits that are found in more than 50% of category members are predictive in that a given category member is more likely than not to exhibit that trait.

“question under discussion” (Roberts, 2006). For example, a speaker might want to make sure the listener avoids mosquitoes, but the listener might be more interested in gauging the true risk so that they can decide for themselves. To get the information they want from a generic statement, listeners need to consider whether speakers are trying to convey a prescriptive or descriptive message. To get their intended message across with a generic statement, speakers need to consider whether listeners are looking for prescriptive or descriptive information.

Kao et al. (2014) provide a model of such a negotiation in the context of non-literal language. If someone says that their coffee pot “cost a million dollars,” the listener cannot reasonably interpret the statement as literal information about the price of a coffee pot. Instead, a listener would interpret the statement as communicating the speaker’s emotional reaction to the price of the coffee pot, namely that the speaker is displeased with how much it cost. The listener can use that information to infer that the coffee pot costs more than what the speaker would consider reasonable, but the main message is how the speaker feels about it. Kao et al. model this behavior as resulting from inferences about the question under discussion. Implicitly, “the coffee pot cost a million dollars” is primarily directed at discussing what the speaker feels rather than what the price is.

Since these two considerations are related, the modeling framework proposed by Kao et al. allows for the possibility that listeners consider multiple questions under discussion; listeners will weight their interpretations according to how likely each option is. For example, if a listener infers that a statement is entirely focused on the speaker’s emotions, the listener can effectively disregard the quoted price when reasoning about the actual price—“a million dollars” has no bearing on the actual price of a coffee pot apart from conveying the speaker’s displeasure. If the speaker quotes a more reasonable price (e.g., “that coffee pot cost, like, two hundred bucks”), the listener might infer that the statement is still meant to convey the speaker’s emotions but is also serving as an approximation of the actual price. Speakers adjust their use of such language based on how they think listeners would interpret a statement, which is in turn based on what listeners are likely to infer about the question under discussion.

Applying this perspective to generic statements points to a potentially important difference between implied prevalence and true prevalence. If the speaker wants to make sure listeners act on the generic statement, then they may be willing to endorse generic statements which imply that the prevalence of a trait is markedly greater than their own conception of the true prevalence. And if listeners first infer that a generic statement is prescriptive, they may then infer that the generic statement's implied prevalence could very well be markedly greater than the true prevalence. We could think of this as an example of the speaker employing hyperbole to make their point, but hyperbolic claims are typically not meant to be taken literally. In this case, speakers may intend for their statement to be taken at face value, since listeners would be all the more likely to adopt the prescribed behavior.

To further examine how speakers and listeners think about the purpose of generic statements, I could continue using *League of Legends* as a naturalistic experimental setting. I would simply provide participants with various generic statements and ask them to rate the extent to which they think the generic statement is meant to inform or prescribe a course of action. I could then examine if speakers' processes for deciding when to make a generic statement, as described by the model outlined in Chapter 4, change as a function of whether they think of the generic statement as prescriptive or descriptive. Since participants may change their assessments of whether a generic statement is prescriptive based on the relative expertise of the speaker and listener, a naturalistic setting like *League of Legends* is particularly useful for providing a complex environment in which people with varying levels of authentic expertise will naturally use generic statements. It is also important to make sure the generic statements are appropriately situated within an authentic culture so that participants can envision the potential purposes of a generic statement.

This line of research could also benefit from parallel studies in a more traditional, tightly-controlled experimental setting. In *League of Legends*, players' goals can be complex and multifaceted. In one sense, it is desirable for people to have different goals, since the whole point is to examine how people's varying goals influence the way they use and learn from generic statements. If everyone had the same goal, this would be impossible. However, players' goals can vary so widely that it becomes difficult to identify exactly why speakers would want to use a generic statement or what listeners would be hoping to get from a generic statement. For example, in a multiplayer game like *League of Legends*, there are social norms within the community of players (Kou & Nardi, 2013). Speakers and listeners will differ in how much emphasis they place on such social considerations, and it is difficult to predict how various social considerations might influence their approach to communicating descriptive information about the environment or prescriptive information about what actions will help players win.

In a more traditional setting, I can control participants' goals by creating an environment with a standardized reward structure (e.g., points). There would need to be some degree of uncertainty about how best to obtain rewards from the environment, so that there is still room for generic statements to provide important information. I would also want participants to have straightforward choices to make, so that I can readily infer how they are using and interpreting such generic statements. Finally, I would want there to be a simple unit for measuring experience in the environment, so that I can readily manipulate participants' expertise and inform them about the expertise of a speaker or listener.

The bandit task, in which participants draw prizes from various gambling machines in search of the best reward, can meet each of these requirements. It can have a standardized reward structure, be it points or money, and participants do not know the true nature of each gambling machine. Participants only need to decide which gambling machine to use on each trial, and I can communicate someone's expertise based on how much experience they have with each gambling machine. In the traditional bandit task, participants need to explore the environment (i.e., sample various machines) so that they can determine the best way to exploit it. In my version, generic statements from a more experienced speaker can serve as a shortcut for the listener, teaching the listener how to increase their rewards without them having to explore the environment themselves.

Unlike more naturalistic experiments, where I can simply ask participants to evaluate whether a generic statement is prescriptive or descriptive, I would want to prompt participants to treat generic statements as prescriptive or descriptive. Since the bandit task both standardizes and simplifies participants' goals, it does not make much sense to ask participants to evaluate whether a generic statement about the environment is descriptive or prescriptive. There may be edge cases where speakers prioritize giving listeners the information they need to make their own decision (e.g., the speaker does not want to assume that the listener shares their own risk propensity), but speakers and listeners are presumably only interested in communicating descriptive information insofar as it helps listeners make decisions that will produce the maximum rewards. If an expert speaker is communicating information about the gambling machines rather than advocating a specific behavioral policy, it probably means that their experiences have been inconclusive. I could prompt participants to think about generic statements in descriptive terms by telling them that their goal is for the next participant to have an accurate understanding of the gambling machines regardless of how many points they get, and I could prompt participants to think about generic statements in prescriptive terms by telling them the opposite.

5.3 Developmental Trajectory of Generic Statements

Neither of the above-mentioned semantic accounts specify how generic statements fit into a learning trajectory. Indeed, typical semantic accounts (e.g., Cohen, 1999; Leslie, 2008; Tessler & Goodman, 2019a; Van Rooij & Schulz, 2020) begin with listeners already having whatever prior knowledge they need to interpret a generic statement and end once they have done so. But how do listeners acquire the prior knowledge they need to interpret generic statements? How do listeners' interpretations change as they gain experience? And how do speakers facilitate the learning process?

5.3.1 Acquiring Prior Knowledge

The most common explanation for how listeners acquire the prior knowledge they need to interpret generic statements is that they transfer their knowledge about similar situations. In particular, many researchers point to what listeners know about how the referenced trait (or a similar trait) is distributed in similar categories (Cohen, 1999; Cimpian et al., 2010; Sterken, 2016; Tessler & Goodman, 2019a; Van Rooij & Schulz, 2020). For example, Cimpian et al. found that listeners tend to interpret generic statements more narrowly if the generic statements reference a trait that would likely occur by accident (e.g., "luzeks are covered in mud") rather than a trait that is likely inherent (e.g., "luzeks have red feathers"). Listeners cannot have prior knowledge about luzeks because they are an imaginary species, but they do know what it means for other animal categories to be covered in mud or have red feathers. Listeners thus expect there to be differences in how strongly those traits are linked to an animal category, even if they have not previously encountered that animal category.

A particularly instructive theoretical framework for thinking about this issue is provided by the idea of overhypotheses. An overhypothesis is broad knowledge that both informs and is informed by more specific knowledge (Goodman, 1983). For example, we have broad knowledge that the color of an animal's feathers is often characteristic of the species as a whole, based on specific examples of species where this is the case. In turn, our specific expectations about a species are informed by our broader expectations about animals in general. Such hierarchical prior knowledge allows us to learn efficiently, since we can both slot individual experiences into a larger conceptual framework and use that larger framework to contextualize individual experiences (Kemp et al., 2007).

Applying this idea to generic statements, overhypotheses can help explain how listeners come to have the prior knowledge they need to interpret a generic statement, even when the generic statement references a topic with which they have no experience. In such a situation, the listener would apply their most relevant overhypothesis. For the generic statements that "luzeks have red feathers," the listener could apply an overhypothesis pertaining to how feather coloring tends to be distributed in feathered animals. Even for a generic statement like "mosquitoes carry malaria," which references a trait that is entirely unique to mosquitoes, a listener who had no direct experience with mosquitoes or malaria could still apply a less directly pertinent overhypothesis about how a similar trait tends to be distributed (i.e., how insects in general tend to act as disease vectors). Such an overhypothesis could be based on, for example, ticks carrying Lyme disease, which would help the listener anticipate that relatively few species members actually carry the disease at any given time.

Overhypotheses can also help us think about how listeners' interpretations change as they gain experience. Additional experience could alter the content, applicability, and weight of listeners' overhypotheses. Considering these possibilities in order, experience might teach listeners that the content of an overhypothesis is incorrect and needs to be adjusted. For example, maybe listeners base their initial assumptions about feather coloring on extensive experience with parrots, where feather colors are unusually diverse. They would initially interpret "luzeks have red feathers" as applying narrowly. Once they realize that parrots are an exception in this regard, they would interpret it as applying more broadly. Experience might also provide listeners with a more directly applicable overhypothesis. Perhaps they initially did not distinguish between feather coloring and any other physical trait, such as body shape. As such, they applied their overhypothesis about how reliably the body shape of a single animal predicts that of their entire species to interpreting both "luzeks have red feathers" and "luzeks have thin beaks." Later, they learn that feather coloring is often more variable, allowing them to distinguish between these two generic statements in terms of how broadly they should be applied. Third, experience might make listeners more confident in an overhypothesis. If they base their overhypothesis on a single example, they may be less committed to it than if it was based on a litany of examples. For example, if a listener thinks any given animal will have the same feather coloring as about 20-30% of its species, the listener will be more open to interpreting "luzeks have red feathers" in a way that conflicts with this overhypothesis (e.g., as implying that 80% of luzeks have red feathers) if it is based on a single example rather than hundreds of examples.

Notably, this line of thinking resembles the implied prevalence account of the semantics of generic statements. Both involve listeners interpreting generic statements by updating prior knowledge. However, I am suggesting that the overhypotheses, on which that prior knowledge is based, change as listeners gain expertise. In contrast, the implied prevalence account suggests that listeners are always reasoning about the prior knowledge of a hypothetical novice listener, regardless of their own expertise. Perhaps people base their understanding of a hypothetical novices' prior expectations on their own overhypotheses, in which case that understanding would change as their own overhypotheses do. However, this possibility is not directly addressed by the implied prevalence account.

Indeed, neither the implied prevalence nor the criteria-based accounts of generic statements distinguish between situations in which speakers are using generic statements to communicate with novices or with fellow experts. Instead, they assume that generic statements are either true or false, with a speaker always willing to say true generic statements and always unwilling to say false generic statements regardless of the expertise of their audience. The implied prevalence account proposes that speakers always evaluate the truth of a generic statement based on whether a hypothetical novice listener would find the generic statement useful. The criteria-based account proposes that speakers always evaluate the truth of a generic statement based on the same set of criteria.

5.3.2 Changing Interpretations

Contrary to such accounts, listeners do seem to change how they interpret generic statements as they gain expertise. Chapter 2 outlines preliminary evidence that this change can be explained in terms of changes to listeners' overhypotheses. In those results, we see experts adjusting their interpretations of generic statements based on context, whereas novices tend to interpret generic statements unsystematically in relation to context. We also see novices tending to interpret generic statements as applying more broadly overall. Yet the difference between experts and novices is relatively minor, considering how novices have no experience with the subject under discussion.

We could readily explain novices being unsystematic in their interpretations by saying that novices are applying an overhypothesis about how strategic considerations tend to play out in games in general, or even computer games in general. Such an overhypothesis might allow them to keep their interpretations within a reasonable range. For example, they might realize that a generic statement about strategic considerations within a computer game is unlikely to apply universally. Otherwise, the game would be rather boring. However, it would not allow them to distinguish between specific examples within the specific game. In contrast, experts may not need to apply overhypotheses if they have direct experience with the referenced team composition. If they lack such direct experience, then they could still apply overhypotheses that allow them to distinguish between specific team compositions. For example, they may know how a specific character tends to behave, which would inform their expectations about how a team composition incorporating that character would behave.

Notably, this way of interpreting the results from Chapter 2 suggests that the nature of the difference between experts' and novices' interpretations may be domain-dependent. Novices' overhypotheses would change as a function of what they would encounter in the process of gaining expertise, which would depend on the specifics of the environment they are learning about. In the case of the *League of Legends* environment, our results suggest that if novices were to learn more, they would discover (1) that the game is less deterministic than they originally thought and (2) that there are important differences between referenced team compositions. As a consequence, they would learn to interpret generic statements as applying more narrowly and alter their interpretations based on the referenced team composition. But in other situations, novices might learn that generic statements apply more broadly than they had originally thought. For example, a novice learning about biology might initially think that organisms with a certain genotype merely *tend* to have a certain phenotype, only to learn that they almost always do.

Since expertise is a complex notion, there may be aspects of the transition from novice to expert beyond overhypotheses that contribute to changes in how people interpret generic statements. In particular, self confidence and perceptions of social authority may play an important role. In Chapter 2, we see evidence that, as novices become experts, they eventually develop the capacity for a qualitatively different form of interpretation that involves largely ignoring the speaker. When an expert speaker is explaining something to a novice listener, the novice listener will tend to take their word for it. When an expert is explaining something to a fellow expert, the expert listener might have sufficient confidence in their own experience to interpret a generic statement as indicating that the expert speaker is irrational or mistaken. Communication between experts might also differ from communication between an expert and a novice because expert speakers can assume that cultural knowledge is common ground between themselves and an expert listener. Expert speakers could then use generic statements to remind expert listeners of that common ground (e.g., "Character A is good in the early game" could be glossed as "*remember*, Character A is good in the early game") or to encourage expert listeners to reconsider the received wisdom (the same generic statement could be glossed as "Character A is *surprisingly* good in the early game"). Expert listeners would then need to adjust their interpretations of such generic statements accordingly.

5.3.3 Facilitating the Learning Process

One nice part of being human is that we do not have to figure everything out for ourselves. We have access to a trove of prepackaged cultural knowledge, assembled through numerous instances of experts passing on what they know to less experienced people. Generic statements are particularly powerful tools for accumulating a shared trove of knowledge because experts can use them to distill their varied experiences into simple messages. Generic statements also help learners leverage that trove of cultural knowledge by providing simple, actionable information that can apply to a variety of circumstances. Yet in Chapter 3, we see that there can be miscommunication when experts use generic statements to share what they know with novices. Furthermore, we do not see evidence of experts customizing their use of generic statements based on the expertise of their audience in Chapter 4.

This lack of adjustment suggests that a novice listeners' perspective may not play a role in the process by which expert speakers' decide when to make a generic statement. As discussed in Chapter 4, it is difficult for someone who has knowledge to accurately consider the perspective of someone who does not. Further research can help determine whether speakers are aware of the differences in how novice and expert listeners would interpret a generic statements, or if this is instead another case of people being subject to the "curse of knowledge" (Camerer et al., 1989). It is also difficult for speakers to adjust their use of language to fit their audience even if they are aware of their audience's perspective. Yet past research indicates that speakers are capable of catering to novices in simple situations (Arvidsson et al., 2022), particularly if they have the information and social motivation that comes with receiving active feedback from the listener (Gann & Barr, 2014; Jucks et al., 2008; Yoon & Brown-Schmidt, 2019). Further research can help determine what sort of cues would prompt experts to account for a novice listeners' perspective and adjust their use of generic statements accordingly.

However, speakers may not feel compelled to adjust their use of generic statements, regardless of the difference between expert and novice interpretations. As discussed in "5.2 Purpose of Generic Statements," a speaker might be focused on altering a listener's behavior rather than communicating descriptive information about the environment, particularly if the listener is a novice. In such a situation, a speaker might be perfectly satisfied with the novice listener interpreting a generic statement differently than they would themselves, even if the listener's resulting conception of the environment is inaccurate in a descriptive sense.

Speakers' satisfaction presumably depends on how novices' conceptions of the environment impact their behavior. In this dissertation, we see situations in which novices are prone to overestimate the applicability of experts' generic statements. While their resulting conception of the environment is technically inaccurate, it would encourage them to lean more heavily into acting on an experts' input, which may be a rational way to deal with uncertainty (Griffiths, Canini, Sanborn, & Navarro, 2019). But it is important to remember that novices may sometimes be prone to underestimation. As discussed in "5.3.2 Changing Interpretations" the difference between how experts and novices interpret generic statements can vary depending on the domain. The costs and benefits resulting from the behavior associated with overestimation or underestimation may also be domain-dependent. These considerations make it difficult to draw broad conclusions about whether speakers neglecting the difference between expert and novice listeners will tend to pose a problem.

5.3.4 Future Research

I can again use both *League of Legends* and more traditional experimental settings to further examine how generic statements relate to developmental trajectories. As discussed in Chapter 1, *League of Legends* actually offers unique opportunities to study the progression from novice to expert on a finer scale. Players are divided into tiers based on their ELO rating, which provides an objective measure of their performance (Elo, 2008). The experiments described were focused on distinguishing between people who were clearly well-acquainted with the topic under discussion and those who were not. My collaborators and I thus recruited participants who were ranked in at least the 27th percentile of players and participants who had no experience with the game. Our sample of "expert" players includes a vast array of expertise, with some of our participants ranked as high as the 95th percentile. With a large enough sample, it would be interesting to study finer gradations of expertise. It would also be interesting to study the players who were excluded from these experiments because they fell into the gap between novice and expert participants (i.e., those who have experience with the game but have not reached the 27th percentile). It can be difficult to recruit players who have only dabbled in *League of Legends* since they are less likely to frequent *League of Legends* forums, but such players could provide insight into the transition from novice to expert.

To test the extent to which overhypotheses are specifically driving developmental differences in how people interpret generic statements, I would want to create experiments in which I directly manipulate participants' prior knowledge. To do so, I would need to recruit participants who had no experience with the topic under discussion so that I can be reasonably certain that their most relevant prior knowledge is whatever information I supply. For such an experiment to be viable, I would also need to be careful about the nature of the information I supply. First, I would want the information to be somewhat vague for all participants so that there is still room for them to learn something from a generic statement. In other words, I would want the information to tell participants what could reasonably happen. I could do so by providing a range representing the narrowest and broadest possible interpretations. A generic statement can then give them further information about what will typically happen, with the information I supply providing context for interpreting the generic statement. Second, I would want to avoid confounding variables by making sure that the information provided to both groups is compatible, in the sense that all versions of the information are describing the same overall environment.

With these constraints in mind, I could construct experiments that examine how the content, specificity, and weight of listeners' prior knowledge impacts their interpretations. To manipulate the content of participants' prior knowledge, I would simply alter the specifics of the ranges I provide. For example, how do participants interpret a generic statement in the context of knowing that a trait could apply to a category 20 to 80% of the time vs. 30 to 70% of the time? To manipulate the specificity of participants' prior knowledge, I would alter what the ranges describe. For example, I might provide some participants with ranges describing what could happen in general and provide other participants with ranges describing what could happen in specific subsets of situations. Finally, I could manipulate the weight of participants' prior knowledge by altering the quantity of observations on which the ranges I provide them are based. For example, I could tell participants that, based on observing 10 vs. 100 examples, I know that a trait applies to a category 20 to 80% of the time.

This proposed experiment can work equally well with *League of Legends* or with a more traditional experimental setting, such as the bandit task discussed above or an experiment that references novel categories (e.g., Chopra et al., 2019). From a novice's perspective, these experimental settings would be largely the same; characters from a video game that they have never played might as well be slot machines or imaginary animals. Even so, there is a ground truth in the *League of Legends* setting, since the referenced entities are observable rather than imaginary. As such, if I provide participants with information which reflects that ground truth, I can compare participants' interpretations to those of actual players with varying levels of authentic expertise, allowing us to pinpoint what sorts of prior knowledge are relevant to interpreting generic statements in a naturalistic context.

Whether speakers are effective in using generic statements to facilitate learning depends on what speakers and listeners are trying to accomplish. As with studying the purpose of generic statements, I would want to control participants' goals, so that I can examine if and when expert speakers' apparent difficulty in considering a novice listener's perspective would end up creating problems in listeners' subsequent behavior. In the context of the bandit task discussed above, this would translate to examining whether a speaker who had considerable experience with the gambling machine would struggle to consider the perspective of a listener who did not, and whether the speaker's statements would consequently lead the listener to make sub-optimal decisions.

5.4 Real-Life Generic Statements

Generic statements are worth studying because they are commonly used to teach people about real-life environments. Esports environments, and the process of learning about them, resemble real life in important ways but are still just game environments. There is ultimately no substitute for studying how generic statements are used in real-life domains, particularly since the patterns described in this dissertation may well be domain dependent. Certain domains are especially worth studying because generic statements play a central role in how people teach each other. In particular, generic statements seem to play a central role in how we pass on cultural beliefs about social categories. They may also play an important role in how we teach each other about science and political issues.

In social contexts, generic statements are important because they seem to be one of the primary means of transmitting beliefs about social essentialism (Rhodes, Leslie, & Tworek, 2012). Essentialism is the belief that there is some true underlying essence that defines a given category, and individual category members are imperfect reflections this essence. Social essentialism refers to such beliefs being applied to social categories (e.g., race, gender, religion). Regardless of whether people explicitly subscribe to such beliefs, they often act as though they do (Gelman, 2003). Doing so is related to stereotyping because such beliefs imply that membership in a social category is an indelible and predictive part of someone's identity (Mandalaywala, Ranger-Murdock, Amodio, & Rhodes, 2019).

People do not apply essentialist beliefs uniformly across all categories. For example, gender tends to be treated as a more essential category than occupation (Mandalaywala et al., 2019). Culture seems to influence which categories are treated as having defining essences and what traits are ascribed to those essences. For example, in Northern Ireland, people tend to view Protestants and Catholics as particularly essential categories, while this is not the case in Israel or in the United States (Smyth et al., 2012; Diesendruck, Goldfein-Elbaz, Rhodes, Gelman, & Neumark, 2013). Even for categories that are commonly treated as essential across cultures, such as gender, the traits ascribed to the defining essences vary widely across cultures (Mandalaywala et al., 2019).

Generic statements are likely an important means of transmitting such cultural information, because they can indicate which categories should be viewed as essential, and they can also specify the features of the defining essence (Rhodes et al., 2012). A statement like "boys are good at math" implies that boys constitute a stable and informative category whose members are reliably characterized as being good at math (Moty & Rhodes, 2021). If a listener already views a category as essential, a generic statement about that category can be interpreted as speaking directly to the nature of the category's defining essence (Prasada, 2000). However, generic statements do not inevitably lead listeners to assume that the referenced category is defined by a central essence; context and prior knowledge are also important to listeners' interpretations (Vasilyeva, Gopnik, & Lombrozo, 2018).

Generic statements are also important to how people teach each other about science and political issues. In scientific contexts (Haigh, Birch, & Pollet, 2020), we often discuss scientific consensus via generic statements (e.g., "scientists believe that gingivitis is linked to heart disease"). If people overestimate how broadly such generic statements apply, they may accept the findings as definitive and fail to investigate further. Alternatively, if people underestimate how broadly such generic statements apply, they may assume that there is considerable disagreement in the scientific community and discount important conclusions. Political statements are also often generic statements, with the applicability of such generic statements being an important part of the message. For example, people across the political spectrum would probably accept that "immigrants make valuable contributions to the economy," but a key point of contention is how broadly this generic statement applies. Since our own prior expectations are important to how we interpret generic statements, we may be prone to talking past each other, agreeing to basic facts but disagreeing on how broadly those facts apply.

To study how generic statements function as pedagogical tools in each of these domains, I could conduct studies similar to those described in this dissertation. I would recruit participants with varying degrees of expertise regarding the topic under discussion and ask them how broadly they believe a generic statement applies, whether they would make the statement themselves, and how they would interpret the statement if a knowledgeable speaker decided to make it. Studying real-life generic statements creates complications, because I lose those simplifying benefits of using esports as an experimental setting that led me to use it in the first place. The real world is messy, particularly when we consider controversial topics. For example, generic statements about vaccine safety and efficacy may now be subject to political and emotional considerations. Such considerations are far more difficult to account for than quantitative estimates of prevalence. However, this is precisely the point of studying such generic statements: real-life topics have unique considerations that can influence how we use generic statements.

5.5 Concluding Thoughts

At its core, this dissertation is about how we teach each other. Generic statements are one of the most fundamental forms of teaching because they involve distilling potentially complex knowledge and experiences down into more easily digestible statements of fact or opinion. In this dissertation, I have pointed to situations in which a learner would interpret a generic statement differently than what a teacher had in mind when making the statement.

This disparity may not actually pose much of a problem, depending on teachers' and learners' goals in using and listening to a generic statement. Furthermore, learners can correct this disparity as they gain experience. In the meantime, though, if learners over- or under-apply what they learn from a generic statement, then they have a misconception about the world that can lead to suboptimal behavior until it is corrected.

I do not mean to suggest that teachers and learners should focus on perfect communication; a disparity will often be inevitable. We would not expect learners to understand the world as their teacher does from a single generic statement. Even if we only consider learners' understanding of the prevalence of a referenced trait, as I have done in this dissertation, inferring that from a generic statement can require extensive prior knowledge that learners may simply lack.

Instead, we can settle for helping learners move in the correct direction. A generic statement can be pedagogically useful so long as it helps a learner develop a more accurate understanding of the world relative to what they previously had, even if that understanding is inaccurate in an absolute sense. It may be that the art of teaching involves guiding learners through a progression of misconceptions, each of which is more useful to the learner than their previous misconception. But in helping learners reach more useful misconceptions, it is worth remembering that a learner will interpret generic statements according to their own prior knowledge rather than that of their teacher.

Bibliography

Arvidsson, C., Pagmar, D., & Uddén, J. (2022). *When did you stop speaking to yourself? Adolescent development of world knowledge-based audience design* (preprint). PsyArXiv. doi: 10.31234/osf.io/mrqgv

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204. doi: 10.1017/S004740450001037X

Bonny, J. W., & Castaneda, L. M. (2016). Impact of the arrangement of game information on recall performance of Multiplayer Online Battle Arena players. *Applied Cognitive Psychology*, *30*(5), 664–671. doi: 10.1002/acp.3234

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, *97*(5), 1232–1254.

Carlson, G. N. (1977). *References to kinds in English.* Ph.D dissertation, University of Massachusetts, Amherst.

Carlson, G. N. (2006). Generics, habituals and iteratives. In K. Brown (Ed.), *The encyclopedia of language and linguistics* (2nd edition ed., p. 18-21). Oxford: Elsevier.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. doi: 10.1016/0010-0285(73)90004-2

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. doi: 10.1207/s15516709cog0502_2

Chopra, S., Tessler, M. H., & Goodman, N. D. (2019). The first crank of the cultural ratchet: Learning and transmitting concepts through language. *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, 226–232.

Cimpian, A., Brandone, A. C., & Gelman, S. A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive Science*, *34*(8), 1452–1482.

Cimpian, A., & Erickson, L. C. (2012). Remembering kinds: New evidence that categories are privileged in children's thinking. *Cognitive Psychology*, *64*(3), 161–185. doi: 10.1016/j.cogpsych.2011.11.002

Cimpian, A., & Park, J. J. (2014). Tell me about pangolins! Evidence that children are motivated to learn about kinds. *Journal of Experimental Psychology: General*, *143*(1), 46–55. doi: 10.1037/a0031687

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in Psychology*, *9*, 287–299.

Cohen, A. (1999). Generics, frequency adverbs, and probability. *Linguistics and Philosophy*,

22, 221–253. doi: 10.1023/A:1005497727784
Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. doi: 10.1016/j.tics.2009.01.005
Dabbish, L., Kraut, R., & Patton, J. (2012). Communication and commitment in an online game team. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 879–888. doi: 10.1145/2207676.2208529
Damen, D. (2018). The Curse of Knowing: The influence of explicit perspective-awareness instructions on perceivers' perspective-taking. *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*, 1578–1583.
Diesendruck, G., Goldfein-Elbaz, R., Rhodes, M., Gelman, S., & Neumark, N. (2013). Cross-cultural differences in children's beliefs about the objectivity of social categories. *Child Development*, *84*(6), 1906–1917. doi: 10.1111/cdev.12108
Elo, A. E. (2008). *The rating of chess players, past and present.* Ishi Press International.
Gann, T. M., & Barr, D. J. (2014). Speaking from experience: audience design as expert performance. *Language, Cognition and Neuroscience*, *29*(6), 744–760. doi: 10.1080/01690965.2011.641388
Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* Oxford University Press.
Gelman, S. A., Coley, J. D., Rosengren, K. S., Hartman, E., Pappas, A., & Keil, F. C. (1998). Beyond labeling: The role of maternal input in the acquisition of richly structured categories. *Monographs of the Society for Research in Child Development*, *63*(1), i–157. doi: 10.2307/1166211
Gelman, S. A., Goetz, P. J., Sarnecka, B. W., & Flukes, J. (2008). Generic language in parent-child conversations. *Language Learning and Development*, *4*(1), 1–31. doi: 10.1080/15475440701542625
Gelman, S. A., Leslie, S.-J., Was, A. M., & Koch, C. M. (2015). Children's interpretations of general quantifiers, specific quantifiers and generics. *Language, Cognition and Neuroscience*, *30*(4), 448–461. doi: 10.1080/23273798.2014.931591
Gelman, S. A., Star, J. R., & Flukes, J. (2002). Children's use of generics in inductive inferences. *Journal of Cognition and Development*, *3*(2), 179–199. doi: 10.1207/S15327647JCD0302_3
Goodman, N. (1983). *Fact, fiction, and forecast.* Harvard University Press.
Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: the case of action video game players. *Cognition*, *101*(1), 217–245. doi: 10.1016/j.cognition.2005.10.004
Griffiths, T., Canini, K., Sanborn, A. N., & Navarro, D. (2019). *Unifying rational models of categorization via the hierarchical Dirichlet process* (preprint). PsyArXiv. Retrieved from `https://osf.io/ketw3`
Haigh, M., Birch, H. A., & Pollet, T. V. (2020). Does 'scientists believe...' imply 'all scientists believe...'? Individual differences in the interpretation of generic news headlines. *Collabra: Psychology*, *6*(17174). doi: 10.1525/collabra.17174
Hawthorne-Madell, D., & Goodman, N. D. (2019). Reasoning about social sources to learn from actions and outcomes. *Decision*, *6*(1), 17–60. doi: 10.1037/dec0000088
Helton, W. S. (2007). Expertise acquisition as sustained learning in humans and other animals: Commonalities across species. *Animal Cognition*, *11*(1), 99–107. doi: 10.1007/

s10071-007-0093-4
Hollander, M. A., Gelman, S. A., & Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology*, *38*(6), 883–894. doi: 10.1037//0012-1649.38.6.883
Jucks, R., Becker, B.-M., & Bromme, R. (2008). Lexical entrainment in written discourse: Is experts' word use adapted to the addressee? *Discourse Processes*, *45*(6), 497–518. doi: 10.1080/01638530802356547
Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*(33), 12002–12007. doi: 10.1073/pnas.1407479111
Kaye, L. K., Pennington, C. R., & McCann, J. J. (2018). Do casual gaming environments evoke stereotype threat? Examining the effects of explicit priming and avatar gender. *Computers in Human Behavior*, *78*, 142–150. doi: 10.1016/j.chb.2017.09.031
Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3). doi: https://doi.org/10.1111/j.1467-7687.2007.00585.x
Kim, J., Keegan, B. C., Park, S., & Oh, A. (2016). The proficiency-congruency dilemma: Virtual team design and performance in multiplayer online games. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4351–4365. doi: 10.1145/2858036.2858464
Kollar, P. (2016). The past, present and future of League of Legends studio Riot Games. *Polygon*.
Kou, Y., & Gui, X. (2014). Playing with strangers: Understanding temporary teams in League of Legends. *Proceedings of the first ACM SIGCHI annual symposium on computer-human interaction in play - CHI PLAY '14*, 161–169. doi: 10.1145/2658537.2658538
Kou, Y., & Nardi, B. (2013). Regulating anti-social behavior on the internet: The example of League of Legends. *iConference 2013 Proceedings*, 616–622.
Kou, Y., & Nardi, B. (2014). Governance in League of Legends: A hybrid system. *Proceedings of the 9th International Conference on the Foundations of Digital Games*, 10.
Kuznekoff, J. H., & Rose, L. M. (2013). Communication in multiplayer gaming: Examining player responses to gender cues. *New Media & Society*, *15*(4), 541–556.
Leavitt, A., Keegan, B. C., & Clark, J. (2016). Ping to win?: Non-verbal communication and team performance in competitive online multiplayer games. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4337–4350. doi: 10.1145/2858036.2858132
Leslie, S.-J. (2008). Generics: Cognition and acquisition. *Philosophical Review*, *117*(1), 1–47. doi: 10.1215/00318108-2007-023
Leslie, S.-J. (2012). Generics articulate default generalizations. *Recherches linguistiques de Vincennes*(41), 25–44. doi: 10.4000/rlv.2048
Leslie, S.-J., Khemlani, S., & Glucksberg, S. (2011). Do all ducks lay eggs? The generic overgeneralization effect. *Journal of Memory and Language*, *65*(1), 15–31. doi: 10.1016/j.jml.2010.12.005
Lindstedt, J. K., & Gray, W. D. (2019). Distinguishing experts from novices by the Mind's Hand and Mind's Eye. *Cognitive Psychology*, *109*, 1–25. doi: 10.1016/j.cogpsych.2018.11.003

Mandalaywala, T. M., Ranger-Murdock, G., Amodio, D. M., & Rhodes, M. (2019). The nature and consequences of essentialist beliefs about race in early childhood. *Child Development*, *90*(4), e437–e453. doi: https://doi.org/10.1111/cdev.13008

Mieg, H. A., & Evetts, J. (2018). Professionalism, science, and expert roles: A social perspective. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (2nd ed., pp. 127–148). Cambridge University Press. doi: 10.1017/9781316480748.009

Milella, V. (2021, November). *League of Legends rank distribution in solo queue - November 2021.*

Moty, K., & Rhodes, M. (2021). The unintended consequences of the things we say: What generic statements communicate to children about unmentioned categories. *Psychological Science*, *32*(2), 189–203. doi: 10.1177/0956797620953132

Pappas, A., & Gelman, S. A. (1998). Generic noun phrases in mother–child conversations. *Journal of Child Language*, *25*(1), 19–33. doi: 10.1017/S0305000997003292

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria*, *124*.

Prasada, S. (2000). Acquiring generic knowledge. *Trends in Cognitive Sciences*, *4*(2), 66–72. doi: 10.1016/S1364-6613(99)01429-1

Ratan, R. A., Fordham, J. A., Leith, A. P., & Williams, D. (2019). Women keep it real: Avatar gender choice in League of Legends. *Cyberpsychology, Behavior, and Social Networking*, *22*(4), 254–257. doi: 10.1089/cyber.2018.0302

Ratan, R. A., Taylor, N., Hogan, J., Kennedy, T., & Williams, D. (2015). Stand by your man: An examination of gender disparity in League of Legends. *Games and Culture*, *10*(5), 438–462. doi: 10.1177/1555412014567228

Reeves, S., Brown, B., & Laurier, E. (2009). Experts at play: Understanding Skilled Expertise. *Games and Culture*, *4*(3), 205–227. doi: 10.1177/1555412009339730

Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, *109*(34), 13526–13531. doi: 10.1073/pnas.1208951109

Richter, S., & Lechner, U. (2009). Transactive memory systems and shared situation awareness: A World of Warcraft experiment. *Proceedings of the International Conference on Organizational Learning, Knowledge and Capabilities*, 1–16.

Roberts, C. (2006). Context in dynamic interpretation. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 197–220). Oxford, UK: Blackwell Publishing Ltd. doi: 10.1002/9780470756959.ch9

Röhlcke, S., Bäcklund, C., Sörman, D. E., & Jonsson, B. (2018). Time on task matters most in video game expertise. *PLOS ONE*, *13*(10), e0206555. doi: 10.1371/journal.pone.0206555

Ruvalcaba, O., Shulze, J., Kim, A., Berzenski, S. R., & Otten, M. P. (2018). Women's experiences in esports: Gendered differences in peer and spectator feedback during competitive video game play. *Journal of Sport and Social Issues*, *42*(4), 295–311. doi: 10.1177/0193723518773287

Seya, Y., & Shinoda, H. (2016). Experience and training of a First Person Shooter (FPS) game can enhance useful field of view, working memory, and reaction time. *International*

Journal of Affective Engineering, *15*(3), 213–222. doi: 10.5057/ijae.IJAE-D-15-00014

Smyth, K., Pendergrast, C., Feeney, A., Coley, J. D., Eidson, R. C., & Niens, U. (2012). The inductive potential of religion categories in northern ireland. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, *34*(34), 2351–2356.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(3), 485–493.

Sterken, R. K. (2016). Generics, covert structure and logical form. *Mind & Language*, *31*(5), 503–529.

Sulik, J., & Lupyan, G. (2018). Perspective taking in a novel signaling task: Effects of world knowledge and contextual constraint. *Journal of Experimental Psychology: General*, *147*(11), 1619–1640. doi: 10.1037/xge0000475

Tardif, T., Gelman, S. A., Fu, X., & Zhu, L. (2012). Acquisition of generic noun phrases in Chinese: learning about lions without an '-s'*. *Journal of Child Language*, *39*(1), 130–161. doi: 10.1017/S0305000910000735

Tessler, M. H., & Goodman, N. D. (2019a). The language of generalization. *Psychological Review*, *126*(3). doi: http://dx.doi.org/10.1037/rev0000142

Tessler, M. H., & Goodman, N. D. (2019b). *Learning from generic language* (preprint). PsyArXiv. Retrieved from `https://osf.io/hnm8p`

Van Rooij, R., & Schulz, K. (2020). Generics and typicality: A bounded rationality approach. *Linguistics and Philosophy*, *43*(1), 83–117.

Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (2015). *Model comparison and the principle of parsimony* (Vol. 1; J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels, Eds.). Oxford University Press. doi: 10.1093/oxfordhb/9780199957996.013.14

Vasilyeva, N., Gopnik, A., & Lombrozo, T. (2018). The development of structural thinking about social categories. *Developmental Psychology*, *54*(9). doi: http://dx.doi.org/10.1037/dev0000555

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189.

Wuertz, J., Bateman, S., & Tang, A. (2017). Why players use pings and annotations in DOTA 2. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1978–2018. doi: 10.1145/3025453.3025967

Yoon, S. O., & Brown-Schmidt, S. (2019). Audience Design in Multiparty Conversation. *Cognitive Science*, *43*(8), e12774. doi: 10.1111/cogs.12774